



A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/78767>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

The Remit of Reasons

by

Joseph John Cunningham

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Philosophy

University of Warwick, Department of Philosophy

September 2015

Contents

Acknowledgements	iii
Declaration	iv
Abstract	v
Introduction	ix
1 Responding to Reasons	1
1.1 Normative Reasons	3
1.2 Responding to Normative Reasons	8
1.3 The Good Case/Bad Case Distinction	14
1.4 Rationalising Explanation	17
2 The Motivating Reasons View	23
2.1 Introducing the Motivating Reasons View	25
2.2 The Dialectic of Psychologism	32
2.2.1 Davidson's View	32
2.2.2 Dancy's View	34
2.2.3 The Neo-Davidsonian View	40
2.3 The Way Ahead	44
3 The Myth of Motivating Reasons	46
3.1 Being Motivated by a Reason as Responding to a Normative Reason . . .	47
3.2 Dancy's Master Argument Rejected	50
3.3 A Dialectical <i>Impasse</i> ?	55
3.4 The Argument from Rationality Rejected	60
3.5 The Argument from Incredulity Rejected	67
3.6 The Argument from Transparency Rejected	68
3.7 The Motivating Reasons View Rejected	73
3.8 Re-Drawing the Map of the Terrain	74
3.8.1 The Dialectic of Psychologism	74

3.8.2	The Primacy of Motivating Reasons	75
3.9	A Merely Linguistic Dispute?	76
4	The Dialectic of Psychologism	78
4.1	Two Arguments for Anti-Psychologism Defended	80
4.1.1	The Linguistic Argument	80
4.1.2	The Counter-Intuitiveness Argument	89
4.2	Two Arguments for Psychologism Rejected	94
4.2.1	The Argument from Error	94
4.2.2	The Argument from Doubt	96
4.3	The Explanatory Question	101
5	Williams' Dictum	106
5.1	The Elements of Success	107
5.2	Williams' Dictum	114
5.3	Reductionism	119
5.3.1	Success for Davidson	120
5.3.2	Success for Dancy	122
5.3.3	Reductionism & Williams' Dicta	124
5.4	The Essential Normativity Thesis	126
5.5	Two Arguments against Williams' Dicta Rejected	132
5.5.1	The Appeal to the Epistemic Thesis	133
5.5.2	The Anscombean Argument	135
5.6	Against Normative Epiphenomenalism	138
5.6.1	Shifting the Burden of Proof	138
5.6.2	Williams' Dicta are Undermotivated	141
5.7	The Primacy of the Good Case	149
	Conclusion	153
	Bibliography	158

Acknowledgements

Many people have helped me in the completion of this thesis, both philosophically and otherwise. I'm not able to remember all of them and as a result what follows is, I'm afraid, only a partial list. My apologies to those who I have missed.

My greatest philosophical debts are to my supervisors at Warwick, Matt Soteriou and Johannes Roessler. Their influence is present on every page and I thank them greatly for their generous support. I would also like to thank Paul Snowdon, who supervised me whilst a graduate student at UCL. Along with many other of his students, he supplied me with what I regard as a model for the way philosophy should be done.

I am grateful to many friends, colleagues and acquaintances for philosophical discussions concerning the material here over the years, too numerous to remember. In particular, I'd like to thank: Ed Nettel, Alex Geddes, Craig French, Barnaby Walker, Wesley Chai, David Rowthorn, Pete Fossey, Nicola Spinelli, Roberta Locatelli, Ivan Ivanov, Alfonso Anaya, Juan Camilo Espejo-Serna, Keith Wilson, Susanne Mantel, Eva Schmidt, Duncan Pritchard, Bill Brewer, Naomi Eilan, Guy Longworth, Matt Nudds, Hemdat Lerman, Tom Crowther, Tim Williamson, Jen Hornsby, Rory Madden, Mike Martin, Mark Kalderon and José Zalabardo.

The work to be presented here has been made possible by the financial assistance of the Arts and Humanities Research Council, UK. I acknowledge and thank them greatly for their support.

Finally, I'd like to thank my parents, Louise and Andrew Cunningham, and grandparents, Kathleen Cunningham and June Southern, for their ever present support and kindness. I'd also like to thank John, Anne and Katie Stray for their encouragement and continual interest. My biggest thanks go to Sarah and Jasmine. The former for all her love, encouragement and assistance. The latter for ensuring that I was up at dawn most days during the last eighteen months of work.

I dedicate the work which follows to the memories of Dennis Southern and Jim Cunningham.

Declaration

I hereby confirm that the following thesis is entirely my own work and has not been submitted as part of a degree at any other university.

Joe Cunningham

September 2015

Abstract

There are reasons for us to act in certain ways and think certain things. We can recognise those reasons for what they are and respond to them accordingly. This thesis is an investigation of that phenomenon. In particular, it constitutes an attempt to resist certain ways in which our understanding of it can be distorted by letting our account of it be informed by *bad cases* of responding to reasons: cases in which one takes something to be a reason and responds accordingly by acting or thinking in the way it seems to one to recommend, even though one fails (blamelessly) to act or think for a genuine reason. I examine and reject three ways in which the possibility of bad cases might distort our thinking about the phenomenon at issue. First, we should reject the thought that the subject is able to act or think for the reason that *p* in both bad cases and good, so that acting or believing for a reason cannot simply be identical to the success condition of responding to a genuine reason. Second, we should reject the thought that the reasons for which we act and think are psychological features of ourselves, because that must be so in bad cases, and what goes for bad cases goes across the board. Finally, we should reject the thought that acting in response to a genuine reason involves only a rationalising explanation of the same type as that which is present in the bad case. The result is a vindication of the position promoted by Raz (2011): responding to reasons is a matter of acting or thinking in a way than manifests one's knowledge of the reasons there are for one to so act or think, so that the reasons in question, which are usually facts about the external world, explain why one does so *qua* reasons. Bad cases are a different kind of thing entirely.

Introduction

One kind of feature of the normative landscape which has received a good deal of philosophical attention in recent times is that of a *reason*. By this, I mean what are usually called *normative reasons*: what I take here to be facts, typically about the external world, which count for or against our performing a certain action or adopting a certain attitude. That the charity does effective work in developing countries is a reason to donate to them, that lighting the candle will cause a fire, given the presence of gas in the room, is a reason not to light it, that my football team have a new manager is a reason to hope that they will be promoted this season, that the weathervane is pointing north is a reason to believe that the wind is blowing in that direction, and so on.

There are many debates that it is possible to have about normative reasons. The debate to be focused on here could properly be described as a debate at the intersection of normative philosophy and the philosophy of mind: what our account should be of what it is to *respond* to normative reasons by acting or thinking in the way they favour. The sort of phenomenon here is commonplace: one recognises that a certain fact favours doing or thinking something and one thereby acts or thinks accordingly, in a way that manifests one's recognition of the facts which favours it. Recognising that the car has a flat tyre and hence that I ought to pull over in order to replace it, I do so and I do so in a way that manifests my knowledge of the normatively significant fact in question: I respond to a reason, in this case a reason for action.

We can contrast the phenomenon of responding to a normative reason with cases of acting or holding an attitude irrationally and a-rationally. Dancing around in jubilation in response to one's team scoring the winning goal, intending to repaint the fence even though one knows one doesn't have any means of doing it, believing that drinking three glasses of red wine a night is healthy out of wishful thinking, and so on, are all cases of either irrational or a-rational activity, which we can contrast with activity that counts as a response to normative reasons.

But we can also contrast the phenomenon of responding to normative reasons with what I call *bad cases* of responding to reasons. A bad case is a case in which the subject takes something to be a reason for action or thought and acts or thinks accordingly, in a way that manifests their taking a fact to be a normative reason for them to so respond, but they blamelessly fail to count as responding to a normative reason nevertheless. This sort of

case contrasts both with the success phenomenon of responding to a normative reason and the non-rational phenomena just described. It contrasts with the latter because unlike the latter, and like the former, the agent is acting or holding an attitude in a way that manifests their taking something to be a normative reason for them to do so. It contrasts with the former however because the agent blamelessly fails to respond to a genuine reason. The most straightforward way in which this can be is if the agent blamelessly gets what reasons there are for action or thought wrong.

Here is how I think we should understand the success phenomenon. Responding to a reason is a matter of engaging in an action or holding an attitude in a way that is subject to a special sort of explanation. The explanation is an explanation why one is performing the action or holding the attitude in question which makes the action or attitude intelligible by appeal to the very facts which speak in favour it – the normative reasons to which one is responding. The act or attitude is not only made intelligible by those facts but is made intelligible by them *in so far as they have the normative dimension they do*, and the sort of explanation at issue essentially requires the subject to know the fact in question as well as to recognise its normative dimension. Here now is how I think we should understand what's going on in the bad case. In the bad case, the subject acts or holds an attitude in a way that is to be explained by appeal to the fact that they believe that p, where p is what they take to be a reason in favour of their so responding. Their action or attitude manifests their taking p to be a reason in favour of their doing it or holding it. That explanation in terms of the subject's belief holds in the good case too, it's just that also, there is a kind of explanation that essentially appeals to the normative reasons present as well. Moreover, I think there is a certain sense in which the explanation in terms of belief is parasitic on the explanation in terms of the normative reason. Throughout, it is taken to be the case that normative reasons are normally facts about the external world. And it is taken that our ordinary talk of acting or believing for reasons can be interpreted as talk which reports the success condition as conceived above, or else as talk which tracks explanations in terms of belief.

The conception of responding to normative reasons just broached is, in a certain way, *minimalist*. It is minimalist in so far as it operates with at most two notions of a reason: the notion of a *normative* reason and the notion of an *explanatory* reason. Normative reasons are facts which exemplify a certain sort of normative dimension; explanatory reasons are items which explain why something is the case. The story just broached says that in cases of responding to genuine reasons we should think of the reason why the agent acts or holds the relevant attitude as identical to the normative reason to which they are responding, where the normative reason explains *qua* normative and the explanation is enabled by the agent's knowledge. Of the bad case we should say that the reason why the agent acts or holds the relevant attitude is identical to the fact that they believe that p, where p is the (apparent) normative reason for their action or attitude.

Perhaps surprisingly, the standard way of conceiving of the phenomena at issue here, which I call the *Motivating Reasons View*, rejects the picture just offered. According to the Motivating Reasons View, we need to allow that the subject can act, believe, desire, intend, judge... for the reason that p in *both* good cases *and* bad. So even if I get it wrong that the car has a flat tyre and act to change it merely because I believe it to have a flat tyre, the proponent of the Motivating Reasons View will say, I still count as acting for the reason that the car has a flat tyre. Since I do not respond to a normative reason in such a case, it cannot be that the concept of a reason in operation is the concept of a normative reason. Rather, it must be a fresh concept of a reason, a concept of something that plays a certain motivating role: a *motivating* reason.

If the Motivating Reasons View is correct, then a number of interesting consequences follow:

- (i) We would have to allow that we have a concept of a reason which is distinct from the concept of a normative reason and the concept of an explanatory reason.
- (ii) We would have to say that the condition of acting or holding an attitude for the reason that p is not simply identical to the success condition of responding to the normative reason that p.
- (iii) We would have to allow that we have a condition which is generic to the good case and bad: the condition for acting or holding an attitude for the reason that p. We would then have to account for what it is to achieve success – what it is to respond to a genuine reason – by saying that it is a matter of having a motivating reason, but where, also, there is a genuine normative reason on the scene which is suitably connected to one's motivating reason. Further, we would have to account for the bad case by appeal to the idea that it is a case in which the subject has a motivating reason, but is missing at least some of the additional factors required for success.
- (iv) The Motivating Reasons View puts pressure on us to subscribe to a *psychologistic* conception of the reasons for which we act or hold attitudes. That's because, whatever motivating reasons are, they are present in the bad case. But also: they explain why the agent acts or holds the relevant attitude. However, what explains must be a fact. The only fact on the scene in bad cases which could do the explanatory work is the fact that the subject believes that p. So the reasons for which one acts or holds attitudes must always be that one believes that p.

As should be apparent, none of (i)-(iv) are consequences of the more minimalist alternative I favour.

This essay is a rejection of the Motivating Reasons View in favour of the more minimalist alternative. I begin, in Chapter One, with an attempt get clear on the basic phenomena

in play here: normative reasons, responding to normative reasons, bad cases of responding to normative reasons, explanatory reasons, rationalising explanations, and so on. In Chapter Two I introduce the Motivating Reasons View as well as the views of Davidson, Dancy, and the Neo-Davidsonians, which are differing versions of the Motivating Reasons View. In Chapter Three I attempt to refute the Motivating Reasons View by showing that there are no such things as motivating reasons in the special sense at issue. Subjects do not act, believe, desire, intend... for reasons in bad cases at all. In Chapter Four I address a well-worn dialectic in this area: whether we should go psychologistic about the reasons which motivate us. In light of my rejection of the Motivating Reasons View, I argue that we should not. Finally, in Chapter Five, I argue for the conception of act and attitude explanation which is at the heart of my more minimal alternative to the Motivating Reasons View.

There is a preliminary point which needs to be made before the enquiry begins. Many of the authors I discuss in what follows focus on the case of acting and omitting for reasons in particular, and ignore what it is to hold an attitude for a reason. Moreover, some authors focus on holding beliefs for reasons and ignore acts, omissions, and non-doxastic attitudes. In contrast, my focus here is *general*: I assume that there is something which can be said about what it is to respond to reasons in general, whether the response takes the form of an act, omission, or the holding of an attitude, whether doxastic or not. I assume that responding to a reason exemplifies a generic structure common to all the different kinds of response there are. Thus, whenever I quote from an author who focuses solely on action or on belief, I assume that what they say can be generalised to anything that can count as a reaction to reasons and proceed on the basis of that assumption.

I do not attempt substantive defence of the assumption that things can be generalised in this way here, but I think it has initial plausibility. However, it should be noted, by way of demonstrating that the assumption is more innocuous than it might at first appear, that I am not claiming that there are simply *no* philosophically interesting differences between acting and believing for reasons or between holding doxastic attitudes and non-doxastic attitudes for reasons. One difference between acting and believing for reasons, for example, is that the former, at least sometimes, involves acting in the pursuit of a goal supplied by a state of mind which either is, or at least mimics, a state of desire. The latter involves no such conative input. That is entirely consistent with the generality assumption operated with here, which is simply that there is a generic structure common to all instances of responding to reasons. The work which follows is an investigation of that structure.

Chapter 1

Responding to Reasons

In the *Groundwork of the Metaphysics of Morals* Kant draws a famous distinction. The distinction is between acting *in accordance* with duty and acting *for the sake of* duty.¹ One acts in accordance with duty if, and only if one's action is prescribed by some moral principle which specifies a duty that applies to one. One acts for the sake of duty, however, only if one performs one's action in the light of its being one's duty to do so.

Kant's distinction is cast in terms of the notion of duty, but an analogous distinction can be drawn with respect to *reasons*. Just as there are actions that are in accordance with the duties that apply to one, so there are actions that are in accordance with the reasons there for one to do it. And just as there are actions performed out of duty, so there are actions which are done for the sake of the reasons for one to act. Acting in accordance with reasons just requires that one's action is favoured by the reasons there are for one to do it. Acting for the sake of reasons requires that one acts in the light of those reasons.

Kant's focus is only on the case of bodily acts and omissions. But the distinction, transposed to the domain of reasons, can be drawn with respect to *mental* acts and omissions that can be done for reasons as well, such as watching, judging, deciding, reasoning, deliberating and concentrating. It can also be drawn with respect to *attitudes* that can be held for reasons such as, for example, believing that p, doubting that p, desiring that p, hoping that p, intending to ψ , regretting that p and fearing that p.

Following Scanlon (1998), let's call any act, omission, or attitude that can constitute a response to reasons *judgement sensitive*.² Let's say that a judgement sensitive expression is any expression that picks out some type of judgement sensitive act, omission or attitude. And let's use ' ϕ ' as a variable for any judgement sensitive expression. Thus, it is possible in general to distinguish ϕ -ing *in accordance with* reasons from ϕ -ing *for the sake of* reasons. ϕ -ing in accordance with reasons is a matter of ϕ -ing in a way that is favoured by the

¹See Kant (1996:4:397-4:398).

²Strictly speaking, Scanlon thinks that the only items which are judgement sensitive are attitudes. Actions are derivatively judgement sensitive: actions can count as responses to reasons only in so far as intentions to act can do so. See Scanlon (1998:21).

reasons there are for one to ϕ . ϕ -ing for the sake of reasons is a matter of ϕ -ing on the basis of reasons.

My concern in this essay is a certain way of thinking about ϕ -ing for reasons which, I think it is fair to say, permeates the contemporary analytic literature. I call the view the *Motivating Reasons View*, for reasons that will become clear in a moment. I lay out the Motivating Reasons View, as well as the versions of it defended by Davidson (1963), Dancy (2000) and others, in the next chapter. But, in broad outline, it is this. The view starts with the thought that we can distinguish *normative reasons* from a perfectly ordinary notion of what motivates the agent to ϕ captured by talk such as: ‘S is ϕ -ing for the reason that p’, and cognate expressions. Normative reasons are facts, typically about the external world, which make it the case that we ought to ϕ . But the View, as I shall put it, identifies the ordinary notion of what motivates with the special notion of a *motivating reason*. Motivating reasons are said to be all and only the entities which have each of the following properties: (i) they are explanatory in the following sense: for there to be a motivating reason for S to ϕ is for S’s ϕ -ing to be subject to a rationalising explanation; (ii) they are reasons; (iii) the rationalising explanation in question is constitutively independent of the existence of genuine normative reasons for S to ϕ , even though it requires the presence of apparent normative reasons for S to ϕ ; (iv) hence motivating reasons and normative reasons are *reasons* in different senses. The notion of a motivating reason, so understood, is then put to work in providing a reductive analysis of what it is to ϕ in response to genuine normative reasons.

The Motivating Reasons View is a generic thesis which comes in many different varieties. Some proponents of the View hold that motivating reasons are psychological states which cause the agent’s ϕ -ing. Others deny that that’s so, and there are other dimensions of difference between competing versions of the View, as we will see. My aim in this essay is to undermine Motivating Reasons Views across the board. With the Motivating Reasons View rejected, this helps to clear the way for a solution to the issue of whether we should think of the reasons for which the subject ϕ s as psychological items. It also enables a more clear minded discussion of the issue of what we should say about success cases of ϕ -ing for reasons. Those two issues are taken up in turn, after the Motivating Reasons View is rejected.

In this chapter I introduce some machinery in the theory of reasons the purpose of which is to enable me to unpack the Motivating Reasons View in the next chapter. By the end of this chapter, the meaning of the theses attributed to the Motivating Reasons View should already have come into clearer focus. I intend my setting up of the terrain here to be non-controversial. Inevitably, some of what I say will be controversial, however, and I flag it up if that’s so. In particular, I make some controversial metaphysical and psychological assumptions. The assumptions are, I hope, both *prima facie* plausible and ultimately defensible, but I don’t undertake a defence of them here: because of limits of

space, that is a project best undertaken on a different occasion.

I proceed in this chapter as follows. In §1.1 I introduce the important phenomenon of a normative reason. In §1.2 I try to pinpoint the phenomenon of ϕ -ing in response to normative reasons. I contrast that phenomenon with a number of other phenomena in order to bring it into clear focus and I describe some of the psychological properties of the subject which are necessary for responding to normative reasons in the sense at issue. In §1.3 I describe the important distinction between good cases and bad cases of ϕ -ing for reasons. Finally, in §1.4 I bring onto the scene the notion of rationalising explanation and relate it to the other notions already on the table.

1.1 Normative Reasons

When we reason with a view to answering the question: ‘Should I ϕ ?’ we typically ask ourselves the further question: ‘What reasons are there for me to ϕ ?’ When we ask this question, we want to know what can be said for and against our ϕ -ing. We want to know what considerations there are in favour of our ϕ -ing and what considerations there are which disfavour our ϕ -ing. Once we have amassed all the considerations we can, we then go on to determine whether or not we should ϕ , all-things-considered, where this typically involves weighing up the considerations for and against. If we decide that the considerations in favour outweigh those against then we ϕ , or else form an intention to do so in the future, if ϕ -ing is the sort of thing with respect to which one can form intentions in the first place.

Considerations which favour and which disfavour one’s ϕ -ing are *normative reasons*. Those are the items we attempt to muster and typically weigh up when we are deciding whether or not to ϕ . They are considerations that stand in a certain normative relation – that of favouring, or disfavouring – to a certain act, omission, or attitude open to one. For reasons of simplicity, in what follows I focus on normative reasons in favour of one’s ϕ -ing and leave out normative reasons with the opposing valency.

That the cat is meowing forlornly at feeding time is a reason in favour of my feeding her, and indeed a reason to think that she wants food. That my football team has reached the play-offs is a reason in favour of my watching the match and if they go on to win, that would be a reason in favour of my celebrating in some way. If however, they lost, then I’d have reason to grumble about it to friends and relatives. Typically, we specify normative reasons using phrases of the form ‘that p is a reason for S to ϕ ’, instead of the more cumbersome ‘that p is a reason in favour of S ’s ϕ -ing’, and I will sometimes follow the common usage in what follows.

I’ve said that reasons stand in the normative relation of favouring to one’s ϕ -ing. To say that they do so is to say little more than that they are reasons in favour of one’s ϕ -ing. That isn’t a problem for me because it is not incumbent on me to give a non-circular

analysis of what it is for a normative reason to be a normative reason here.³ But even though I do not offer a non-circular analysis of the notion of a normative reason, I do want to offer an informative characterisation of it in what follows. I do not offer a complete characterisation of the notion, but I hope to say enough to make it clear enough for my purposes.

Favouring is a normative relation, as I have said. When a certain consideration stands in favour of one's ϕ -ing, we can apply certain normative predicates to one's ϕ -ing, or at least to the prospect of one's ϕ -ing, if one has not already engaged in ϕ -ing yet.⁴ We can say that one would be justified in ϕ -ing – that one would be justified in adopting the relevant attitude or performing the relevant action.⁵ We can say that one ought to ϕ – that one ought to adopt the attitude, or perform the action.⁶ And we can say that it would be sensible, appropriate, and reasonable for one to ϕ .⁷ In what follows I restrict my attention to the status of its being the case that one ought to ϕ . I do so only for the sake of simplicity, it would make little difference were I to focus on any of the other normative statuses mentioned instead.

Most normative reasons provide defeasible support for one's ϕ -ing. They make it the case that one *pro tanto* ought to ϕ : they make it so that one ought to ϕ , *as far as the relevant consideration holds*. Whether those reasons make it so that one ought to ϕ *all-things-considered* or *overall* depends on what countervailing reasons there are and what the strength of those countervailing reasons are. When there is a set of reasons against ϕ -ing which are stronger when taken together than the set of reasons for ϕ -ing, ϕ -ing is not something one ought to do, *all-things-considered*. One's reasons for ϕ -ing have been defeated by one's reasons against ϕ -ing. Nevertheless, one's reasons for ϕ -ing continue to make it so that one ought to ϕ *pro tanto*. They continue, even when defeated by stronger countervailing reasons, to favour one's ϕ -ing: it continues to be the case that those considerations constitute something to be said in favour of ϕ -ing. We can call those types of normative reasons *pro tanto* reasons.

Most, if not all, reasons for mental and bodily action are *pro tanto*. That it's raining outside is a reason for me to take my umbrella, but that reason can be outweighed by countervailing reasons – for example, that my umbrella is cumbersome and that it is prone to fold in the wind – such that *all-things-considered* I ought not take my umbrella. In

³For the record, I don't think it is possible to offer such a non-circular analysis. The notion of a reason in favour is best treated as primitive. For a defence of that idea see, for example, Scanlon (1998:Ch.1) and Dancy (2004a).

⁴What follows is not an exhaustive list.

⁵The notion of justification in operation here needn't be thought of as the only notion of justification there is. For example, it might be that there is a further notion of justification we possess which a particular ϕ -ing can instantiate only when one has ϕ ed, or is in the process of ϕ -ing, for a normative reason. That second type of justification isn't guaranteed to be present just by dint of a normative reason being present.

⁶This will only be plausible if it's correct that 'ought' needn't always mean, and perhaps never means 'is obligated to' and if it's correct that 'ought' needn't always mean, and perhaps never means 'morally ought'. For these points see Alvarez (2010:11) and Broome (2004). I assume here that each of those claims about the meaning of 'ought' is correct.

⁷Those seem to me to be different ways of saying the same thing.

those circumstances, the fact that it is raining still means that there's something to be said in favour of performing the action which, overall, I ought not perform. Similarly, that the music is beautiful is a reason for me to attend to it. But that reason might be outweighed by counterveiling reasons – for example, that I need to concentrate on something more pressing – such that I ought not do it all-things-considered. Some reasons for belief are pro tanto too. That there are fresh tyre tracks impressed into the mud on the drive-way is a reason to believe that someone has just left in the car. But it can be outweighed by counterveiling reasons for belief – for example, that the car is currently out of service – such that it is not the case that all-things-considered I ought believe that someone has just left in the car. Still, in those circumstance, the fact about the fresh tyre tracks continues to count in favour of my believing that someone has just left in the car.

By contrast, some normative reasons provide indefeasible support for one's ϕ -ing, at least in the sense that it is not possible for them to make it the case that one ought to ϕ at all, and yet it not be the case that all-things-considered one ought to ϕ . We can call those types of normative reasons *indefeasible reasons*. A good deal of reasons for belief are indefeasible. That the man in the room is playing a middle C on the piano is an indefeasible reason for me to believe that he's not playing a G. That there is a pig in the field I'm looking at is an indefeasible reason for me to believe that there is a mammal in the field. That my friend Stephen is married is an indefeasible reason for me to believe that he isn't a bachelor. In each case, we have a consideration that can obtain only if p. Because of that, it is not possible for that fact to make it the case that one ought to ϕ at all, and yet be outweighed by counterveiling epistemic reasons.⁸

When a certain normative reason makes it the case that one ought to ϕ all-things-considered, we can say that that reason decides the issue of whether one ought to ϕ . Thus, we can label reasons which have that status *decisive reasons*. Given what's been said, decisive reasons come in two forms: they can come in the form of pro-tanto reasons which are not, relative to a particular context, outweighed by counterveiling considerations and they can come in the form of indefeasible reasons. All facts that constitute indefeasible reasons are decisive reasons, in whatever context; indefeasible reasons simply guarantee that one ought to ϕ all-things-considered, if they make it the case that one ought to ϕ at all. But a pro-tanto reason can be a decisive reason in one context, when there are no stronger counterveiling reasons on the scene, and not a decisive reason in others, when there are such stronger counterveiling reasons present.⁹

⁸It might be the case that moral reasons in general are indefeasible in contexts where there are no counterveiling moral reasons present, because the presence of a moral reason 'silences' all counterveiling non-moral reasons in the sense that the presence of the moral reason is sufficient to extinguish the normative force of the counterveiling non-moral reasons. That claim is, of course, highly contentious. For a defence of it, see McDowell (1979).

⁹I've been talking as if whenever we have a case of pro-tanto reasons having the status of decisive reasons that that is a matter of there being a single consideration that constitutes a pro-tanto reason for one to ϕ and which is not outweighed by counterveiling considerations. I will continue to talk as if that's so, in order to

I've said enough about normative reasons, I hope, to have pinpointed the phenomenon adequately enough. Now I turn to the following question: what are normative reasons? That it's going to rain is a reason in favour of my taking an umbrella with me when I go outside. That the exit-poll predicts a Labour victory at the by-election is a reason in favour of my believing that Labour will win. But to which category should we assign normative reasons? My answer to this question is, I think, the orthodox one: reasons in favour of one's ϕ -ing are best thought of as facts. It is *the fact* that it's going to rain which is my practical reason and it is *the fact* about the exit-poll which is my epistemic reason. This is one of the controversial assumptions I make. It is worth codifying:

(NR=F) All normative reasons are facts.

(NR=F) says that for anything that's a normative reason, that thing is a fact. It does not imply that all facts are normative reasons, although that might well be so.¹⁰

What are facts? One option is that they are states of affairs understood as coarsely individuated entities composed of particulars and properties which constitute the unification of those particulars and properties.¹¹ A second option is they are obtaining states of affairs understood as coarsely individuated abstract entities that can exist whilst failing to obtain.¹² A third option is that they are truths: true propositions understood as finely individuated entities composed of Fregean senses that constitute the content of thoughts and which are ways of thinking about concrete reality.¹³ On any of those options, it looks as if facts, and hence, given (NR=F), normative reasons, are not concrete objects, events or property instances. Is that a correct thing to say about facts? I assume so here – I assume that reasons are facts, that facts are either states of affairs (on some conception thereof) or truths, and that this rules it out that facts are concrete particulars.¹⁴

Given (NR=F), normative reasons are mind-independent in the sense that they are facts, and whether or not a certain fact obtains is a mind-independent affair. That is not to say all or even some facts that constitute normative reasons have their normative status in a mind-independent way – it is not to say that the normativity of normative reasons is 'part of the fabric of the world', to borrow the influential phrase from Mackie (1977). Consistently with (NR=F), it could be that normative reasons for action have their normative status

keep matters simple. But it should be noted that perhaps the typical case involves not a single consideration having the status of a decisive reason but a set of considerations where *together* those considerations are decisive, though taken individually they are not.

¹⁰It would be so, for example, if all facts were reasons to believe themselves.

¹¹See Amrstrong (1997) for a defence of that conception of facts.

¹²See Textor (2012) for an exploration of states of affairs on that conception.

¹³See Dodd (2008) for a defence of that third conception of facts.

¹⁴Because there is more than one normative reason, given (NR=F) I'm committed to thinking that there is more than one fact. That is yet another claim I take for granted, contentious though it is. Davidson (1969) offers a Slingshot Argument for the claim that if there are such things as facts, then there is only numerically one fact. Part of what I'm taking for granted is that there is an adequate way to respond to Davidson's Slingshot.

because of what one does or would take a pro-attitude towards, in ideal conditions of information and reasoning, for example.¹⁵ The thought is only that given (NR=F), normative reasons are mind-independent in the sense that the entities which are normative reasons enjoy a mind-independent existence. That implies nothing about the mind-independence of the normative status of normative reasons.

As well as displaying that minimal level of mind-independence, there is a minimal sense in which normative reasons display objectivity. That is, there is a minimal sense in which subjects can be wrong about what normative reasons there are. That can happen in either of two ways. First, a subject might make a mistake about what facts there are. They might believe that p , and take p to be a normative reason for them to ϕ , even though it is not the case that p . I might take it that the road up ahead is closed and regard that as a reason to travel in a different direction, even though I am mistaken about what reason I have because I'm mistaken about the road being closed. Second, a subject might get it right that p is a fact, but make a mistake about the normative status of the fact that p . They correctly believe that p , but incorrectly take p to be a reason in favour of their ϕ -ing. I might correctly take it that the exit-poll predicts a Labour victory at the by-election, but mistakenly think that the particular exit-poll is reliable and so mistakenly take the fact that the exit-poll predicts a Labour victory as a reason in favour my believing that Labour will win.

Not only can a subject be wrong about what reasons there are in these two ways, they can be ignorant of what reasons there are in analogous ways too. A subject might be ignorant of the existence of a reason in favour of their ϕ -ing because they are ignorant of the obtaining of the fact that constitutes the normative reason. And a subject might be ignorant of the existence of a normative reason because they are ignorant of the normative status of the fact in question.

We can say, whenever an agent takes p to be a fact and a reason in favour of their ϕ -ing, that is *appears* to them as if p is a reason in favour of their ϕ -ing. It is important to note that *all that's meant* by talk of the appearance of reasons here is that the subject takes p to be a fact and takes it to have the relevant sort of normative significance, whatever those conditions, in turn, amount to. To say that it appears to the agent as if p is a reason, in the relevant sense, is not to say that the subject is in some phenomenally conscious state or is undergoing some phenomenally conscious event, the phenomenal character of which involves p being present to consciousness as a reason. It is rather to say that the subject is in a state, or set of states, which constitutes them taking p to be a fact and to be a reason in favour of their ϕ -ing. Those states needn't involve the presence of any phenomenally conscious mental entity.

When an agent correctly takes p to be a normative reason for their ϕ -ing, we can say

¹⁵See Williams (1980) for the classic contemporary defence of that 'internalist' position about normative reasons for action.

that it appears to them as if p is a reason in favour of their ϕ -ing and does so correctly: p does have the status of a normative reason for them to ϕ . But when an agent mistakenly takes p to be a normative reason for their ϕ -ing we can say that p is a *merely apparent* reason for them to ϕ . Merely apparent normative reasons are not normative reasons, they just appear to subjects to be so even though they are not. Merely apparent normative reasons are common in everyday life, because it is common for people in everyday life to make mistakes about the existence of facts and their normative statuses. But they are also common in the outlandish cases discussed by philosophers. The subject who is a brain-in-a-vat suffering a series of hallucinatory experiences they cannot tell apart by introspection from genuine perceptions is a subject who thinks that there are many normative reasons applicable to them on the scene, even though they are systematically mistaken. Most, if not all considerations taken to be normative reasons by brains-in-vats are merely apparent normative reasons.

A final point about normative reasons is in order. Later on, we'll be encountering the proponent of *Psychologism*, who says, of the considerations on the basis of which the agent ϕ s, that those considerations are all psychological items, for example psychological states of the agent. I take it to be obvious that a Psychologistic position about normative reasons is not an option. Normative reasons, if there are any at all, are often facts about the mind-independent environment. It is the fact that the exit-poll predicts a Labour victory – not in any relevant sense a mental or psychological fact – that's a reason for me to believe that Labour will win. It's the fact that it's going to rain – a fact about the weather – that's a reason for me to take my umbrella, and so on. Along with many others writing about these issues, I'm going to take it for granted that normative reasons needn't be, and often aren't, psychological items, but are often facts about the external world. As should be clear from the caveats included in the formulation of that last point, I am not denying that normative reasons are sometimes psychological items – specifically facts about the subject's psychological states. That I believe I've been abducted by aliens is a reason to visit a psychiatrist, for example. The thought is just that this needn't be the case, and typically isn't.¹⁶

1.2 Responding to Normative Reasons

Let us suppose that an agent decides that there is sufficient normative reason for them to ϕ . For example, let us suppose that an agent determines that there is a certain normative

¹⁶It should be noted that what has been said here is consistent with the thought that there might be certain psychological conditions which need to be fulfilled for a certain fact to count as a reason, whether epistemic or practical. It might be that I need to have certain pro-attitudes in order for the fact about the weather to be a practical reason for me. It might be that I need to know the fact about the exit-poll in order for it to make it the case that I ought to believe that Labour will win and hence be a normative reason for me to believe so. That's entirely consistent with saying that the things which are practical and epistemic reasons needn't be, and typically aren't, psychological items.

reason for them to ϕ , p, and a certain normative reason against them ϕ -ing, q, and they decide that p is decisive. The agent might then ϕ *in response to* the decisive normative reason in question.¹⁷ Thus, the agent would count as ϕ -ing *on the basis of, on the grounds that, or in the light of* the normative reason that p. The fact that p would be the reason *for which* the agent ϕ s.

The sense in which we ϕ *in response to* normative reasons at issue here is that we ϕ in a way that is a manifestation of our recognition that the relevant facts are reasons in favour of our ϕ -ing. The sort of link that holds between one's ϕ -ing and the fact which favours one's ϕ -ing is a special sort of link which holds only when one's ϕ -ing is a response to normative reasons *as such*.

I intend the phenomenon at issue here to be the analogue of the Kantian notion of acting for the sake of duty, transposed to the domain of reasons and generalised to judgement sensitive phenomena across the board. One way of getting a grip on the notion ϕ -ing in response to normative reasons, then, is to compare it to the Kantian notion of acting in the light of duty. Acting in the light of duty, for Kant, should be contrasted with acting merely in accordance with duty. The latter amounts to acting in a way that is in-keeping with the duties that apply to one in the particular context. The action might be in-keeping with duty, but that might well be an accidental feature of the action. Actions performed *for the sake of duty*, by contrast, are actions which are not only performed in-line with duty but are performed in a way that manifests the agent's awareness of the duty conferring factor as such, so that the action is non-accidentally in-line with duty. The notion of ϕ -ing in response to normative reasons as such should be understood analogously. ϕ -ing in response to normative reasons is a matter, not just of ϕ -ing in accordance with the reasons there are for one to ϕ , but is also a matter of ϕ -ing in a way that manifests one's awareness of the normative reasons as such, so that one's ϕ -ing is non-accidentally in-line with what the reasons recommend.

In order to bring the phenomenon into clearer focus, I now want to contrast it with a number of distinct phenomena which can be described using similar language.

First, we can contrast the sense in which one ϕ s in response to the normative facts which I'm focusing on here with the sense in which the movement of the leaves on a tree can be said to be a response to the gust of wind, the sense in which the perspiration of the man can be said to be a response to the high temperature, and the sense in which the death of the zebra can be said to be a response to dehydration. In each case, what we have is one non-agential natural event being caused by a second event fitting the same description. These are clearly not cases of ϕ -ing in response to reasons as such – they do not even qualify as ϕ -ings in the first place, given that ' ϕ ' ranges over judgement sensitive phenomena.

Second, we can contrast the sense in which one ϕ s in response to the normative facts

¹⁷I stick to a case where there is only one reason on either side purely for the sake of simplicity.

at issue here with the sense in which a neurotically held desire to harm one's friend can be said to be a response to the repressed emotion of jealousy it is controlled by, the sense in which a belief that the Conservatives are competent on economic matters could be said to be a response to the subject's upbringing, and the sense in which jumping around for joy can be said to be a response to the feeling of elation it is controlled by. In each case, what we have is a judgement sensitive phenomenon: a desire, a belief, and an action respectively, but which is in some sense controlled by a condition which is not itself a reason for them to them to respond in the relevant way. As a result, the agents in question cannot be said to be responding to reasons in the sense at issue here, for there are no reasons on the scene to which they could be said to be responding, and in any case the sort of link between their ϕ -ing and the relevant condition is not of the right sort – it is more akin to the sort of mere causal relation holding in the non-agential cases already mentioned.¹⁸

Third, we can contrast the sense in which one ϕ s in response to normative reasons at issue with the sense in which a non-rational occurrent desire to get off the aeroplane can be said to be a response to the moderate turbulence being suffered on board, the sense in which the chicken-sexer's belief that the chick is male can be said to be a response to the fact, sub-personally detected by the subject, that the chick is emitting a type of pheromone that only male chicks emit, and the sense in which one's non-rationally omitting to follow one's friends on their climb up a steep mountain can be said to be a response to one's feeling nervous about the climb. In each case, what we have is a judgement sensitive phenomenon: a desire, a belief and an omission respectively, which is again controlled in some sense by a certain condition, but where this time the condition happens to be a reason in favour of the agent responding in the way they do. As a result, their ϕ -ing is in conformity with the reasons there are for them to ϕ , but they do not ϕ in response to those reasons as such. The connection between their ϕ -ing and the consideration which happens to be a reason in favour of it is not of the right sort.

There's a final contrast worth drawing: a contrast with cases of *weakness of will*. As I'll understand them, these are cases in which the subject recognises that all-things-considered they ought not ϕ , but they ϕ anyway. As such, I understand weakness of will in an undemanding way, so that it doesn't just extend to practical judgement sensitive phenomena like actions, omissions, desires, intentions, hopes and wishes, but extends also to cases of cognitive judgement sensitive phenomena such as believing, judging, doubting and suspension of judgement. The cases of wanton agency discussed by Frankfurt (1971), in which the unwilling addict goes against his better judgement and takes a hit anyway, provide us with an example of the former sort of weakness of will case. The case of the woman who judges, in response to evidence recognised by her to be decisive, that degrees

¹⁸It is controversial, but I think true, that jumping around for joy in a way that's controlled by a feeling of elation one is undergoing is not a case of ϕ -ing in responses to reasons. See Hursthouse (1991) for a compelling defence of that claim.

from foreign institutions are of equal value to degrees awarded by home institutions but who nevertheless goes on to prejudicially believe that foreign degrees are inferior anyway discussed by Peacocke (1998) provide us with an example of the latter sort of weakness of will case.

It might seem as if one could ϕ in a weak-willed way where one's ϕ -ing is also an instance of ϕ -ing in response to normative reasons as such. For consider, in closer detail, some of the features of Frankfurt's case of wanton agency. The unwilling addict recognises full well that they ought, all things considered, to refrain from taking the drug. Nevertheless, they continue to take the fact that taking the drug will bring them pleasure as a reason to take it, just a defeated one. They take the drug anyway and in so doing they 'give in' to their desire for the pleasure of their drug use. Isn't the correct description of this case that the agent takes the drug *for the reason* that doing so will bring them pleasure? And hence, shouldn't we conclude that cases of weakness of will can at least sometimes involve ϕ -ing in response to normative reasons?

In fact, I think it's a mistake to think of the agent in Frankfurt's case as taking the drug in response to a normative reason as such. For the agent recognises that the normative force of the fact that taking the drug will give them pleasure is overridden by countervailing considerations – for example, that it will continue to do damage to their health and that it will enable their addiction to persist, with all of its undesirable ramifications. How, then, could it be that the agent counts as taking the drug *in so far as the fact that it will give him pleasure favours his doing it*? The normative force of the fact which he in some sense or other responds to seems irrelevant to his decision to do it, given that he recognises it to be defeated. The correct thing to say is rather that a non-rational occurrent desire for the drug motivates him to take the drug, in a way analogous to the way in which jumping for joy is non-rationally controlled by one's feeling of joy. Weakness of will, understood as a phenomenon which can extend to judgement sensitive phenomena generally, provides us with a further contrast case with the phenomenon of responding to reasons.¹⁹

I've said enough now to have brought the phenomenon at issue in this section into focus. I take it for granted that we can ϕ in response to facts which favour our ϕ -ing, in the sense at issue. I can believe that labour will win the by-election in response to the fact that the exit-poll predicts a Labour victory. I can take my umbrella in response to the fact that it's going to rain. I can celebrate in response to the fact that my team won, and so on. In each case we have an instance of rationally responding to the facts: an instance of ϕ -ing in response to the relevant fact *qua* normative reason. Subjects who ϕ in response to normative reasons are subjects who ϕ for *good* reasons, in at least one sense of that phrase. In the remainder of this section I want to bring into focus some of what is going on with the subject at the psychological level when they ϕ in response to a normative reason. In particular, there are two psychological features of agents when they ϕ in response to

¹⁹It should be acknowledged that what I have said about the case of wanton agency is, of course, contentious.

normative reasons to which I want to draw attention.

First, one can ϕ in response to the normative reason that p only if the fact that p is present to one's mind in some way. By 'present to mind' I don't mean that p is the object or content of some phenomenally conscious mental state or occurrence. It is consistent with p being present to mind in the sense at issue that p is the object or content of some standing state one is in. Rather, the thought is that ϕ -ing in response to the normative reason that p requires that there is some suitable psychological link between oneself and p. One cannot ϕ for the reason that p if the fact that p is in existence and yet one has no idea about it.

I am going to be operating here with the assumption that the psychological state that constitutes the presence to mind of one's reason when one responds to a reason is a state of belief. Responding to the normative reason that p, that is, requires believing that p, and it is that belief which constitutes the relevant psychological link between oneself and one's reason. We can label this *the Doxastic Thesis*:

(DT) The psychological state that constitutes the presence to mind of p in the way required for ϕ -ing in response to the normative reason that p is belief that p.

(DT) is a controversial thesis because it rules it out that non-doxastic and more generally non-representational states of mind can do the work of constituting the presence to mind of reasons. It rules it out, for example, that perceptual experiences, thought of as contentful states or not, can constitute the presence to mind of reasons for perceptual belief. (DT) has the status of a simplifying assumption I make throughout this essay. It doesn't do any work in any of the arguments I develop later on. If one is sympathetic to the idea that non-doxastic or non-representational states can do the relevant work then (DT) may be substituted for the preferred, more liberal principle.

(DT) doesn't say that the *only* property of the relevant representational state which enables it to play the role of constituting the presence to mind of reasons in the way required for responding to reasons as such is the property of its being a belief that p. It could be that the relevant belief needs to have further epistemic properties, such as being rationally held, justifiably held or being an instance of knowledge, if it is to play the role at issue. This is significant, because in Chapter Five I will be arguing that nothing short of knowledge is indeed required for responding to normative reasons as such.²⁰

There is a second claim about the psychology of ϕ -ing in response to normative reasons which it will be useful to have on the table. I call it the *Normative Thesis*:

²⁰It is also worth noting that (DT) is consistent with the thought that the state which constitutes the presence to mind of normative reasons when one ϕ s in response to them is the state of knowing, *but where the state of knowing is conceived of as a primitive, factive state which cannot be partly reduced to believing*. That is the influential conception of knowledge one finds developed by Williamson (2000). (DT) simply says that the relevant state of mind falls under the kind *belief*. It is consistent with that that it also falls under the kind *knowledge*, and we are not to conceive of the latter kind as partly reducing to the former.

(NT) Whenever a subject ϕ s in response to the normative reason that p, the subject takes p to favour their ϕ -ing.

(NT) says, in other words, that a subject can count as ϕ -ing in response to the normative reason that p only if p counts, from the point of view of the subject, as a reason in favour of their ϕ -ing. I can count as believing that the Labour Party will win the by-election in response to the normative reason that the exit-poll predicts a Labour victory only if I take the exit-poll prediction as a reason in favour of believing that Labour will win, for example.

I take (NT) to be highly plausible, once it is recognised that what's at issue here is an analogue of the Kantian notion of ϕ -ing for the sake of duty. Just as one can count as ϕ -ing for the sake of duty only if one takes the consideration which provides one with one's duty – for example, that I promised to ϕ – as providing one with a duty to ϕ , so one can ϕ in response to a normative reason only if one takes the consideration that constitutes the normative reason in favour of one's ϕ -ing to have that normative dimension. ϕ -ing for the sake of duty requires taking the duty providing consideration as a duty providing consideration, and that once we transpose things to the realm of reasons, that fact carries over. That point is especially plausible once it's recalled that ϕ -ing for the sake of duty requires ϕ -ing in the light of the duty conferring factor *qua duty conferring factor* – it requires ϕ -ing for the sake of a duty conferring factor *as such*. That implies that one can ϕ for the sake of duty only if one takes the duty conferring factor as a duty conferring factor. Transposing to the case of reasons, an analogous point holds.

It should be noted, for the purposes of making sure that the meaning of (NT) is precisely in view, that the sort of psychological property at issue should be distinguished from that of merely *supposing* that p is a reason in favour of one's ϕ -ing. When offering someone advice about how to go about achieving some goal that they think p is a reason for, one might doubt whether p is really a reason for them, or anyone, to pursue the goal in question, but nevertheless *suppose* that it is for the purposes of the conversation, so that one can then go on to offer one's interlocutor advice about how to achieve the end they have in view. In such cases one isn't committed to the normative status of p, but there is a sense in which one *treats* p as having the normative status in question and there might even be a sense in which one *takes* p to have that status. The sense in which one takes or treats p to be a normative reason in such cases should be distinguished from the sense in which p is taken to be normative when one ϕ s in response to the normative reason that p, according to (NT). According to (NT), what's required is not merely that one supposes p to be normative, but that one is, in some sense, *committed* to the normative status of p.

It might be thought that the only way to read (NT) is as requiring the presence of a further representational mental state, in addition to the state of believing that constitutes the presence to mind of p, which represents the normative dimension of p. Indeed, it might be thought further, given what was said in the preceding paragraph, that the representational state in question must be a further belief in the truth of the proposition that p is a reason

in favour of one ϕ -ing. And one might now worry that (NT) is implausible because it places too much of a demanding restriction on ϕ -ing in response to the reason that p. One reason for thinking the restriction too demanding is because the sort of normative concepts that form part of the content of the representational state in question are concepts that not everyone who is capable of ϕ -ing for reasons possesses – the concept of favouring or of making it the case that one ought to ϕ . But perhaps we can avoid this worry, if indeed it is a real worry, by interpreting what it is to take a certain item as a normative reason in terms of the subject having a certain set of dispositions to respond to that item in certain ways – for example, dispositions to ϕ or be motivated to ϕ when one recognises that p and one doesn't recognise any stronger reason against ϕ -ing, being disposed to cite p as a reason when asked why one should think ϕ -ing a reasonable thing to do, and so on – and then suggesting further that one's having that set of dispositions is insufficient for one to be in a state which represents the normative dimension of p.

There is, then, an issue about whether to interpret (NT) as requiring the presence of a further representational mental state, in addition to the belief that p present when one ϕ s in response to the normative reason that p, or not. I'm going to leave aside that issue here. I will not be saying much about how to interpret (NT) in what follows. I take it for granted that there is some reading of it such that it comes out as a plausible thesis. I leave a further exploration of (NT) and the issues it generates for another occasion.

Given (NT), whenever one responds to a normative reason as such by ϕ -ing, one takes the reason as a normative reason for one's ϕ -ing. But one can take something to be a normative reason without responding to it as such. One does that when one decides that a pro-tanto reason one has in view is outweighed by a reason with the opposing valency. The psychological condition specified by (NT), then, is required for responding to reasons but it doesn't suffice for it.

1.3 The Good Case/Bad Case Distinction

So far, we have on the table the thought that there are such things as facts which favour our ϕ -ing and the idea that we can ϕ in response to such facts, in a sense of 'in response to' which differs the sense in which events in nature are responses to their causes, the sense in which non-rational ϕ -ings are responses to the mental states and events which bring them about, and the sense in which non-rational ϕ -ings are responses to factors which just so happen to be normative reasons in favour of the subject ϕ -ing. We've also looked at some of the psychological features which are on the scene whenever one ϕ s in response to normative reasons too: (DT) and (NT) are the salient theses in that regard.

What I have been focusing on so far, however, is what I will now start calling the *good case* of ϕ -ing for reasons. Good cases are cases in which the agent ascertains the existence of a certain normatively significant fact and rationally responds to it by ϕ -ing, so that their

ϕ -ing can be said to be a manifestation of the awareness of the normative reason as such. But there are also what I will call *bad cases* of ϕ -ing for reasons, which the philosopher interested in this sort of phenomenon should pay attention to, which are relevant to the formulation of the Motivating Reason View in the next chapter, and which I therefore now want to explore.²¹

A bad case of ϕ -ing for reasons, as I'll be understanding it here, is a case in which the psychological features – (DT) and (NT) – associated with good cases hold, so that the subject takes there to be a reason in favour of their ϕ -ing, and in which the subject's ϕ -ing can be said to be a manifestation of their taking p to be a reason in favour of their ϕ -ing, but where they nevertheless blamelessly fail to count as responding to normative reasons as such. Such a subject would, were they to reflect as carefully and as honestly as possible, come to the conclusion that they are ϕ -ing in response to a normative reason, even though they are not. In that sense, their situation ϕ -ing-for-reasons-wise is indistinguishable from their own point of view from the situation of the subject in a corresponding good case, but they do not count as ϕ -ing in response to the normatively significant facts nevertheless.

The most common sort of bad case discussed in the literature is the type of bad case in which the agent blamelessly gets what normative reasons there are for them to ϕ wrong, but ϕ s in a way that manifests their taking p to be a normative reason anyway. This in turn can happen in either, or both, of two ways. First, the subject might blamelessly but falsely believe that p , so that they get what facts there are wrong. Second, the subject might blamelessly but incorrectly take p to favour their ϕ -ing (even though p might be a reason for something else), so that they get the normative status of the relevant facts wrong.

Suppose I decide to water the neighbours' plants in response to the fact that I promised to whilst they're away. But let's suppose, furthermore, that I've blamelessly misremembered: I didn't promise to water the *neighbours'* plants, but the plants of the couple living across the road from me. So it's false that I promised to water the neighbours' plants whilst they're away. But I go to water the neighbour's plants anyway, under the influence of the blamelessly false belief that I promised to do so for them. That would be an instance of a bad case in which I incorrectly but blamelessly take p to be a reason in favour of my ϕ -ing by dint of blamelessly but falsely believing that p . Now consider a case in which I know that the exit-poll predicts a Labour victory. I might believe that Labour will win because I believe that the exit-poll predicts a Labour victory. But, unknowable to me, the exit-poll in question is entirely unreliable and hence the fact about the exit-poll prediction isn't really a reason for me to believe that Labour will win at all. That would be an instance of a bad case in which I incorrectly but blamelessly take p to be a reason in favour of my ϕ -ing by dint of blamelessly but incorrectly taking the fact that p to favour my ϕ -ing.

²¹The good case/bad case terminology is borrowed from Williamson (2000), who applies the labels to the normal epistemic case and the case of the subject who is a victim of a sceptical scenario respectively. The labels have since become common currency in epistemology and the philosophy of perception.

But there are other sorts of bad cases possible. These are cases in which the agent gets their normative reason right, but nevertheless blamelessly fails to ϕ in a way that is linked to the relevant reason in the right sort of way to count as responding to normative reasons as such. The sorts of bad cases at issue include cases in which the subject blamelessly finds themselves in a position where their belief that p fails to have the sort of epistemic credentials necessary for ϕ -ing in response to a normative reason.

Since the subject in the bad case satisfies both (DT) and (NT), we can say that it appears to them as if p is a reason in favour of their ϕ -ing, and that p has, relative to the subject in question, the status of an apparent normative reason. In the first sort of bad case p has the status of a *merely* apparent normative reason. In the second sort of bad case the apparent normative reason is a genuine normative reason, but one doesn't count as responding to it because one isn't hooked-up to the reason in the right sort of way.

In the previous section I contrasted ϕ -ing in response to the normative reason that p with various sorts of cases which can be described using similar language as well as with weakness of will cases. Bad cases of ϕ -ing for reasons also contrast with all of those cases whilst contrasting with cases of genuinely ϕ -ing for normative reasons as well. We can assimilate them to cases of ϕ -ing in response to normative reasons with respect to being cases where the subject ends up ϕ -ing in a way that manifests their taking something to be a reason for them to ϕ . As a result, bad cases differ from the contrast cases discussed above. However, unlike in good cases, the subject fails to genuinely respond to the reasons there are. So bad cases also contrast with good cases.

A final comment is in order before moving on. The psychological features laid down by (DT) and (NT) as being required for responding to reasons are features present in both the good case and the bad case. As such, in both the good case and the bad case it appears to the subject as if p is a reason in favour of their ϕ -ing. It's just that in the good case the appearance is genuine and the subject succeeds in responding to reasons as such, whereas in the bad case either the appearance happens to be genuine or it doesn't, but either way the subject blamelessly fails to ϕ in response to the normative facts. So what we should say is that there is a *generic condition*: ϕ -ing in response to an apparent normative reason, where that condition admits of a successful realisation: ϕ -ing in response to a normative reason – the phenomenon discussed at length above – and a defective realisation: taking something to be a normative reason and ϕ -ing accordingly but blamelessly failing to ϕ in response to normative reasons.²²

²²Of course, there's no suggestion here that those are the most fundamental ways of describing the generic, success and failure notions in this area. It might be that a positive theory of each sort of case is possible.

1.4 Rationalising Explanation

Now I want to bring into the picture the notion of an *explanation why* an event occurred or a fact obtains. The discussion of that phenomenon will enable me, in turn, to introduce a certain type of explanation why to which judgement sensitive entities are uniquely subject: *rationalising explanation*. I shall have much more to say about rationalising explanations in Chapter Five, but some introductory remarks need to be made at this stage because the Motivating Reasons View to be introduced in the next chapter involves theses about the nature of rationalising explanation.

The practice of providing explanations is familiar. We can explain to others *how* to do something – for example how to ride a unicycle; we can explain to others *what* something is – for example what a computer is; and we can explain *why* certain phenomena occur or why certain facts obtain. In what follows I restrict my attention to explanations why.

There is a phenomenon in this area which we also use the language of reasons to talk about but which should be distinguished from the phenomenon of a normative reason: the *reason why* a certain event occurs or fact obtains. These items we can call *explanatory reasons* and they are the items that constitute the *explanantia* of explanations why certain phenomena occur or certain facts obtain. Those sorts of entities count as reasons not because of a normative relation that they stand in to anything, but by standing in an explanatory relation – whatever that amounts to exactly – to what they explain.

Paradigmatically, we offer explanations why by engaging in distinctive illocutionary acts of explaining, using sentences of the following mutually translatable forms:

- (i) p because q
- (ii) What explains why p is that q
- (iii) The reason why p is that q

For example: ‘the man disappeared because he was kidnapped’, ‘what explains why the man disappeared is that he was kidnapped’, and ‘the reason why the man disappeared is that he was kidnapped’. What ‘p’ picks out in each of the above sentences is what is getting explained: the *explanandum* of the explanation being provided. The phrase that we replace ‘p’ with is the *explanandum phrase*. What ‘q’ picks out in each of the above sentences is what does the explaining: the *explanans* of the explanation being provided. The phrase which replaces ‘q’ is the *explanans phrase*. Explanatory reasons are what get referred to by *explanans* phrases in true explanatory sentences.

We’ve already seen that the phrase ‘reason for’ can sometimes pick out the normative relation that normative reasons stand in to ϕ -ings, as in the sentence: ‘a reason for him to take his umbrella is that it is raining’. To this we can now add a second use. Sometimes it singles out the relation of explanation, the relation that links explanatory reasons with the

items that they explain, as in the sentence: ‘the reason for the man’s disappearance is that he has been kidnapped’. When I use the phrase ‘reason for’ in what follows it should be clear which use I’m relying on.

It seems clear that we can only provide true explanations of why a certain event occurs or why a certain fact obtains if the event really has occurred or if the fact really does obtain. Thus, it is true that he has blue eyes because both his parents have blue eyes only if it is in turn true that he has blue eyes. Genuine *explananda* must have occurred or be the case. It has often been thought plausible that the following analogous claim is true about *explanantia*:

The Factivity of Explanation. If p functions as an *explanans* then p is true.

The Factivity of Explanation says that p can explain why a certain event has occurred, or why something is the case, only if p is true. Only when it is a fact that p can we have a genuine, that is: true, explanation on our hands. Again, this looks like a plausible principle and I will be operating with it throughout this essay. As we will see, Dancy (2000) denies it, but he doesn’t say anything in favour of its denial other than that it follows from his own theory, and, as we shall see again, his own theory is objectionable anyway. I should note in passing, however, that it is consistent with denying the Factivity of Explanation that one accepts that some ways of providing explanations require the truth of the *explanans* phrase. It’s consistent with denying Factivity, for example, that the ‘because p’ form and its cognates require the truth of p, even though that is only a contingent linguistic feature of that form, and is not a reflection of the nature of explanation why in general. As we will see, that is the position of Dancy (*ibid.*) himself.

I now want to introduce an important species of explanation why: rationalising explanation. We can provide explanations of natural phenomena. We can explain why the apple fell from the tree at the velocity at which it did by appeal to the fact that it has a certain weight and that it fell from a certain height. We can also provide explanations why of certain agential phenomena, that is: judgement-sensitive phenomena, and there are various ways we can do so. We can explain why someone is holding their knee and whimpering by appeal to the fact that they are experiencing terrible cramp in their knee. That would be to explain a certain action by appeal to a bodily sensation which is functioning as a mere cause of their behaviour. We can explain why the woman is disposed to treat people with kindness by appeal to the fact that she was brought up that way, or by appeal to the fact that she belongs to a culture which places special importance on that virtuous character trait. The first would be to provide an explanation of the trait by appeal to the person’s history. The second would be to explain the trait sociologically. We can explain why the man believes he is being stalked by appeal to the fact that he suffers from paranoid schizophrenia. That would be to offer a medical explanation of the agential phenomena – the man’s belief – in question. And supposing there is some neural process underpinning the para-

noid schizophrenic subject's delusion we might explain why they have it by appeal to the relevant neural process. That would be to provide a neural explanation of the subject's belief.

One important way in which we can give explanations of judgement-sensitive phenomena, distinct from those sorts of explanation just considered, is to explain why the agent ϕ ed by appealing to what was, from the agent's point of view, a normative reason for ϕ -ing. This is the sort of explanation Davidson (1963) introduces here:

What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action. (*Ibid.*:3)

The explanations in question I will call *rationalising explanations* and they are what Davidson refers to here as 'rationalizations'. The *explanantia* of rationalising explanations I will sometimes call, following Davidson, *rationalisers*. Rationalising explanations purport to render the agent's ϕ -ing intelligible by appealing, in some way or other, to what was, from the agent's own point of view, a reason in favour of their ϕ -ing:

A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action – some feature, consequence, or aspect of the action the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable. We cannot explain why someone did what he did simply by saying the particular action appealed to him; we must indicate what it was about the action which so appealed. (*Ibid.*)

Quite how these rationalising explanations work, in particular: with what we should identify rationalisers, and how many kinds of rationalising explanation there are, is a matter of controversy, as we'll see. But I want to make a few remarks about rationalising explanation and the connection between that concept and good cases and bad cases of ϕ -ing for reasons now, so that the Motivating Reasons View can be brought into focus in the next chapter.

One thing to note is that rationalising explanations are a species of explanations why. As such, whatever properties are exemplified by explanations why generally must be exemplified by rationalising explanations in particular. Thus, if the Factivity of Explanation applies to explanation why generally, then it applies to rationalising explanations too. Similarly, if all explanations why are causal explanations, then rationalising explanations are too.

An agent's ϕ -ing is subject to a rationalising explanation just in case there is something which counts in favour of ϕ -ing from the agent's point of view and the agent is ϕ -ing

accordingly. Indeed, that last claim is trivially true, once it's recognised that ' ϕ -ing accordingly' is an explanatory notion. There is something which counts in favour of ϕ -ing from the agent's point of view and the agent ϕ s accordingly in both the good case and the bad case: they satisfy the psychological features laid down by (DT) and (NT) and ϕ accordingly across both cases, after all. So an agent's ϕ -ing is subject to a rationalising explanation in both the bad case and the good case, which is not to say that S's ϕ -ing is subject to a rationalising explanation of only a single form across both cases.

Although both sorts of case involve the presence of rationalising explanations, there are some differences between the good case and bad case which need to be flagged up, rationalising explanation-wise, nevertheless.

When the agent responds to a normative reason – when they are in the good case – their ϕ -ing is subject to a rationalising explanation which we can report using a phrase of the form: ' $S \phi$ ed, or is ϕ -ing, because p '. There is, as we can put it, a non-psychologistic rationalising statement that's applicable to them. In the bad case, the agent's ϕ -ing is subject to a rationalising explanation which we can report using a phrase of the form: ' $S \phi$ ed because they believe that p '. There is, as we can put it, a psychologistic rationalising statement that's applicable to them.

In the bad case, we cannot use a non-psychologistic rationalising statement to provide a rationalising explanation of the agent's ϕ -ing. That's because that sort of phrase can be true only if p is true, given the factivity the 'because' idiom, only if p is a reason in favour of the agent's ϕ -ing and only if S is related to p in the right sort of way, but bad cases are cases in which at least one of those conditions fails, even though it seems otherwise to the subject. In the good case, however, we can use a psychologistic rationalising statement to provide a rationalising explanation of the agent's ϕ -ing, it's just that we can do so also using a non-psychologistic rationalising statement too. So the correct thing to say is that psychologistic rationalising statements are generic. In the good case we can also use non-psychologistic rationalising statements. In the bad case we can only use psychologistic rationalising explanatory statements.

To draw this distinction between two ways of stating rationalising explanations is not, it should be noted, to imply that there are two forms rationalising explanation can take, so that an explanation reported non-psychologically is non-identical to an explanation reported psychologically. On the contrary, it is consistent with what I've said here that there is only one sort of rationalising explanation, and that sort of explanation is reported using each type of sentence, so that whatever differences there are between asserting and endorsing each sentence is extra to the sentences status as reporting a rationalising explanation of that type.

Labelling the sort of explanatory statements we can make only in the good case 'non-psychologistic' is potentially misleading in a way that's worth flagging-up. I noted in §1.1 that although normative reasons are typically facts about the external world, sometimes

normative reasons are facts about one's own psychological states, and in particular facts about what beliefs one has. For example, the fact that one believes that one has been abducted by aliens is a reason in favour of visiting a psychiatrist. One can respond to such psychology-involving normative reasons in the sense at issue. If one does so, we can offer a rationalising explanation of why one ϕ s by saying: 'S ϕ s because they believe that p'. By my lights, that would, contrary to appearances, count as a non-psychologistic rationalising statement. It counts as such because non-psychologistic rationalising statements, as they are to be understood here, contain, at least on the surface, *explanans* phrases which pick out the facts which are the normative reasons to which the subject is responding. Psychologistic rationalising statements by contrast contain, at least on the surface, *explanans* phrases which pick out the psychological state of believing which constitutes the presence to mind of the normative reason to which the subject is responding. So when the subject's normative reason is a psychology-involving fact, the correct form of the relevant psychologistic rationalising statement is: 'S ϕ s because they believe that they believe that p'. In sum, then, non-psychologistic rationalising statements can be statements which, on the surface, include an *explanans* phrase that pick out a psychological fact about the subject who ϕ s, just so long as the fact is the normative reason the subject is responding to, and not the state of belief which constitutes the presence to mind of the relevant normative reason.

Normally, practical judgement sensitive phenomena such as actions, omissions, intentions, and desires are thought to be subject to rationalising explanations stateable in the psychologistic form which cite not just the agent's belief that p but also some pro-attitude the agent has. If that's right, then psychologistic rationalising statements shouldn't say: 'S is ϕ -ing because they believe that p', but: 'S is ϕ -ing because they believe that p and have [e.g.] a desire to ψ ', where ' ϕ ' is restricted to practical judgement sensitive phenomena. An alternative to this view is that such psychologistic rationalising statements can make reference not to some conative pro-attitude the subject has, but to a further belief the subject has, to the effect that p is a reason in favour of their ϕ -ing. Whatever the truth is in this area I ignore the supposed extra components of psychologically stateable rationalising explanations in this essay and focus solely on the contribution made to such explanations by the agent's believing that p. To that extent, the psychologically stateable rationalising explanations I focus on here are always to be understood as incomplete explanations, strictly speaking. An analogous point holds for the non-psychologically stateable rationalising explanations too: insofar as there is an element of them corresponding to a relevant pro-attitude belonging to the subject, that element will be bracketed off throughout this discussion so that only the element corresponding to the subject's belief will be focused on.

The thought that we have in mind, in the good case, when we think of an agent as ϕ -ing because p, is a way of thinking about the phenomenon of ϕ -ing in response to a normative

reason as such. That is to say: recognising a normative reason and responding accordingly just is the state of affairs we have in mind when we think whatever thought it is that gets expressed by a statement of the form ‘S ϕ ed/is ϕ -ing because p’. Studying what we have in mind when we offer rationalising explanations in the non-psychologistic form is, then, a way of getting at the phenomenon of responding to normative reasons as such. What this brings it out is that ϕ -ing in response to the normative reason that p constitutively involves ϕ -ing in a way that is subject to a rationalising explanation reportable using a non-psychologistic rationalising statement.

Likewise, a way of getting at what’s generic to the good case and bad case – taking p to be a normative reason for one to ϕ and ϕ -ing accordingly – is by studying what we have in mind when we offer rationalising explanations in the psychologistic form, for the thought expressed there is a way of thinking about that generic condition. What this brings out is that that generic condition constitutively involves ϕ -ing in a way that is subject to a rationalising explanation reportable using a psychologistic rationalising statement.

We can of course use the phrase ‘S ϕ s because p’ to offer explanations which are not rationalising. We do so when we say, for example, of a paranoid subject S, that S believes that he is being spied on by the man across the road because the man across the road made eye-contact with him. In order to make it clear that what I have in mind when I use one of those phrases is a rationalising explanation I shall sometimes say that I’m using the rationalising sense of ‘because’. This way of talking should be taken with a pinch of salt. I intend it merely to flag-up that the sort of explanation at issue is a rationalising explanation. I do not intend the way of talking in question to commit me to the substantive thesis that the English word ‘because’ expresses different concepts depending on whether it is used in the context of providing a rationalising explanation or in the context of providing a non-rationalising explanation.

This concludes my discussion of the preliminary material we need to have on the table. The next chapter is dedicated to a presentation of the view I want to undermine: the Motivating Reasons View.

Chapter 2

The Motivating Reasons View

In his seminal paper *Actions, Reasons, and Causes* Donald Davidson, it is typically thought, put on the map a theory of ϕ -ing for reasons which, I don't think it is overstating things to say, constituted the norm amongst Philosophers of Action, Philosophers of Mind, Meta-ethicists, and Epistemologists for a generation. I will be following the standard interpretation of Davidson's view here. According to it, the reason for which the subject ϕ s is to be identified with the rationaliser of the their ϕ -ing. Rationalisers are in turn to be identified with certain environment-independent psychological states of the subject which cause their ϕ -ing: suitable pairs of pro-attitudes and beliefs which constitute ϕ -ing being cast in a positive light for the agent. In a word, one's reasons for ϕ -ing are certain psychological states which cause one's ϕ -ing.

In response to what he thought to be the overly psychologistic tendency of that Davidsonian view, Jonathan Dancy, in his book *Practical Reality*, defended an alternative view. According the alternative, the reasons for which the subject ϕ s are to be identified with what the agent takes to be the normative reason for their ϕ -ing, so that, in the good case, the reason for which the subject ϕ s *is* their normative reason. It was then agreed by Dancy that the reasons which motivate the agent to ϕ are to be identified with the agent's rationalisers and it was concluded, given that the reason for which they ϕ is that p even in the bad case, that rationalising explanations needn't have true *explanantia* and hence aren't causal.

Attempting to preserve what they saw to be a correct emphasis on the non-psychological character of the reasons for which the agent ϕ s, but wishing to preserve the factivity and perhaps also the causal character of rationalising explanation as the Davidsonian View does, there were then a raft of Neo-Davidsonian theories defended in response to Dancy, recommended by philosophers such as Smith (2003), Davis (2003, 2005), Olson and Svensson (2005), Setiya (2011), and Hieronymi (2011). What the Neo-Davidsonians have in common is the thought that we should reject an assumption which both Davidson and Dancy share: that the reasons for which the agent ϕ s are the agent's rationalisers. With that assumption rejected, we can coherently agree with Dancy about the reasons

which motivate the agent whilst agreeing with Davidson about the agent's rationalisers.¹

Despite their differences, there is a style of thinking about ϕ -ing for reasons which is common to Davidson, Dancy and the Neo-Davidsonians. All three theories are instances of a generic conception of ϕ -ing for reasons which is presupposed but developed in different ways by each of them. The generic view in question is the Motivating Reasons View, which we've already briefly encountered. It can be introduced in the following way. We have a perfectly ordinary notion of the reason which motivates the agent to ϕ . It is that ordinary notion which gets exercised in thought expressible using sentences such as: 'S is ϕ -ing for the reason that p' and cognate sentences. The Motivating Reasons View begins with an innocuous distinction between the notion of a normative reason and the ordinary notion just pointed to. The View then, quite controversially, goes on to impose a certain theoretical interpretation on that latter notion. It introduces the special notion of a *motivating reason*, which is defined in terms of four distinctive characteristics, most notably of which is that the items in question are present, counting as reasons and doing their motivating work across both the good and bad cases. It identifies the ordinary notion gestured to with the notion of a motivating reason, and goes on to offer a Reductive account of what it is to respond to normative reasons by appeal to it.

This essay argues for a rejection of the Motivating Reasons View. In the next chapter I aim to show that the ordinary notion of a reason which motivates should not be identified with the special notion of a motivating reason because the latter notion is empty: there are no entities which exemplify each of the properties in terms of which the notion is defined. On top of that, it is argued, in Chapter Five, that we ought not provide a Reductive account of what is going on in the good case by appeal to what is common to good and bad cases. The result is that the Motivating Reasons View is rejected root and branch. All of that is yet to come. The purpose of this chapter is merely to *introduce* the Motivating Reasons View, as well as the central versions of it already mentioned, so that the reader has an initial conception of my target at the outset of the enquiry.²

I will proceed as follows. In §2.1 I introduce the Motivating Reasons View. In §2.2 I present the views of Davidson, Dancy, and the Neo-Davidsonians, by way of illustrating the Motivating Reasons View and by way of demonstrating that my target is real. Introducing

¹There was also a less interesting response to Dancy which consisted in defending a more thorough-going Davidsonianism. See, for example, Wallace (2003), in the case of action, and Turri (2009, 2011) in the case of belief.

²It should be made explicit, as has already been intimated, that I am very much not alone in my rejection of the Motivating Reasons View. So-called Disjunctivists about ϕ -ing for reasons, like Hornsby (2007a, 2007b, 2008), McDowell (2013a) and Roessler (2014) should be read as rejecting the Motivating Reasons View as should Hyman (1999, 2006, 2010, 2011), Stout (2004, 2009), Alvarez (2010), Raz (2011), and Marcus (2012). None of those mentioned explicitly use the phrase 'the Motivating Reasons View' for a claim they want to attack, and nor do many of them make it explicit that a target they have in mind is the view for which I'm electing to use that label, but I think they should be read as rejecting the view in question, nevertheless. For at least some of those authors, I make explicit as I go along how I think the claims and arguments to be defended here by way of rejecting the Motivating Reasons View improve on their own.

those views also serves to introduce a familiar dialectic in this area to which Chapter Four is dedicated. Finally, in §2.3 I provide an overview of what is to follow, with the intention of helping to orientate the reader.

2.1 Introducing the Motivating Reasons View

We can begin with a familiar and perfectly correct point. It can be true that the fact that p is a reason in favour of $S \phi$ -ing, even though each of the following sentences, which we can safely assume to have the same meaning, is false of S :

- (1a) $S \phi$ ed/is ϕ -ing for the reason that p
- (1b) The reason for which $S \phi$ ed/is ϕ -ing is that p
- (2a) $S \phi$ ed/is ϕ -ing on the basis of p
- (2b) The basis on which $S \phi$ ed/is ϕ -ing is that p
- (3a) $S \phi$ ed/is ϕ -ing on the grounds that p
- (3b) The grounds on which $S \phi$ ed/is ϕ -ing is that p

The contrast here is real: the fact that p might be a normative reason for one to ϕ without it being the case that one ϕ s for the reason that p , on the basis of p , or on the grounds that p . That the cat is hungry is a reason for me to feed her, but because I haven't realised that she's hungry I fail to act at all and hence fail to act for the reason that she's hungry, for example. There being a normative reason for one to ϕ can come apart from ϕ -ing in a way that is motivated by the reason that p .

What this brings out is that there is a perfectly innocuous distinction to be drawn between two notions. On the one hand, there is the notion of a normative reason, which I've already introduced. When we think that p is a normative reason in favour of $S \phi$ -ing, we're thinking of p as a fact which has a certain normative dimension. On the other hand, there is the notion of ϕ -ing for the reason that p . When we think that S is ϕ -ing for the reason that, on the basis of, or on the grounds that p , what we're thinking is that S is ϕ -ing in a way that is motivated by something else to which we're applying a concept of a reason. To apply that notion just is to think of the relevant entity as a reason which plays the sort of motivational role at issue. That second notion, of a reason that motivates, is paradigmatically exercised in thoughts that have the content expressed by sentences (1)-(3). What's been brought out is that we can have the first thought without having the second, and coherently so. We can think of something as a normative reason without thinking of the agent as being motivated to ϕ by anything, including the very thing to which we're applying the concept of a normative reason.

There is a quite natural way of interpreting sentences (1)-(3) which, ultimately, I think is correct. The natural way of reading them says that for S to ϕ for the reason that p is just for there to be a normative reason which favours S's ϕ -ing, to which S has responded or is responding by ϕ -ing. So understood, (1)-(3) pick out the success condition of ϕ -ing in response to normative reasons which was introduced in the last chapter. The truth of (1)-(3) would then require whatever is required for S to count as responding to a normative reason – for example, that p is true.³ And understood in that way, the concept of a reason which figures in (1a/b) – and which I think is also expressed by the phrases 'basis' and 'grounds' in (2a/b) and (3a/b) respectively – is identical to the concept of a normative reason. The sense in which S is said to have ϕ ed or be in the process of ϕ -ing for *the reason* that p by (1a/b), and by the cognate expressions in (2a/b) and (3a/b), is that they have ϕ ed or are ϕ -ing in response to a *reason in favour* of their ϕ -ing.

If we interpret (1)-(3) in that supposedly natural way, then we will have to say that there being a normative reason for S to ϕ can come apart from ϕ -ing motivated by the reason that p, picked out by (1)-(3), just because a normative reason could exist and have its normative status without the agent responding to it *qua* normative reason. We shall also have to conceive of the relation between the notion of a normative reason and the notion of a reason which motivates the agent in the following way. The notion of a normative reason is the notion of a fact that favours one's ϕ -ing, whereas the notion of a reason that motivates one to ϕ is the notion of such a normatively significant fact playing a certain sort of role: that of being the normative reason to which one's ϕ -ing is a response, in the sense introduced in the last chapter.

The Motivating Reasons View does *not* interpret sentences (1)-(3) in that supposedly natural way. It is argued, by the proponent of the View, that the notion of a reason which motivates the agent to ϕ is the notion of an entity which can be present, and playing the sort of motivating role at issue, across good and bad cases: the subject ϕ s for the reason that p in both good cases and bad. It follows that the item in question is not a normative reason, and that the motivational role at issue doesn't require it to be. Instead, the proponent of the Motivating Reasons View says that the item to which a concept of a reason is applied by sentences (1)-(3) really is a *reason*, to be sure, yet the sense in which it's a reason is not that it's normative, but that it's motivating: the relevant sort of motivating role individuates the notion of reason-hood which is in play when we think of the agent as ϕ -ing motivated by a reason. The item in question is thus said to be a *motivating reason*. Interpreted in that way, the truth of (1)-(3) would not require the subject to count as responding to a normative reason. (1)-(3) could be true in the bad case as well as the good case. Moreover, such expressions deploy a distinctive sense of the word 'reason'. The proponent of the Motivating Reasons View will then go on to give a positive account

³This is how Alvarez (2010), Stout (2004, 2009), Raz (2011) and Roessler (2014) would read (1)-(3), for example.

of what motivating reasons in that special sense are, part of which will be an account of what the motivating role amounts to, and they will offer a Reductive account of the success condition of responding to a normative reason which appeals to the special notion of a motivating reason.

On the Motivating Reasons View we shall have to say that there being a normative reason for S to ϕ can come apart from the phenomenon captured by (1)-(3) because the fact that p could be a normative reason without the agent having the motivating reason that p *and* because one could have the motivating reason that p without the fact that p being a normative reason. On the supposedly natural way of interpreting (1)-(3), by contrast, only the first conjunct corresponds to a real possibility. Moreover, we shall have to interpret the contrast drawn between the notion of a normative reason and the notion of a reason which motivates the agent to ϕ in terms of a contrast between two notions of a reason: those items which have a status of a reason individuated by the normative role they play and those items which have a status of a reason individuated by the motivating role they play. On the supposedly natural way of drawing the contrast, only the first individuates a genuine notion of a reason for us: the latter items are reasons, but they are reasons only in the sense of being normative reasons which, extra to their status as reasons, play a motivating role.

What's just been offered is a relatively inchoate specification of what the Motivating Reasons View amounts to. To sharpen our conception of the View I want to begin by connecting the phenomenon of ϕ -ing in a way that's motivated by the reason that p with the notion of rationalising explanation, introduced in the last chapter. To ϕ in a way that's motivated by the reason that p is a matter of ϕ -ing subject to a rationalising explanation which appeals, in some way, to the reason for which the agent ϕ s. In other words, to ϕ in a way that's motivated by the reason that p consists in ϕ -ing in a way that is subject to an explanation which can be expressed using a sentence containing a rationalising 'because', a sort of explanation which requires that one satisfies the psychological conditions laid down by (DT) and (NT), explored in the last chapter, but where the *explanans* of that explanation either is, or in some way involves, the item which is characterised as the reason for which one ϕ s.

That we should cash out the notion of a reason which motivates in terms of the notion of a rationalising explanation is clear once it is recognised that sentences (1)-(3) are ways of reporting rationalising explanations. In uttering a sentence such as (1), for example, one will have thereby provided one's interlocutor with a rationalising explanation of why S has ϕ ed or is ϕ -ing. They are ways of reporting rationalising explanations which differ, at least at the surface level, from rationalising explanatory statements which use the 'because' idiom, such as the psychologistic and non-psychologistic rationalising statements discussed at the end of the last chapter, but they do report rationalising explanations nevertheless – rationalising explanations which could just as easily be stated using a 'because' idiom.

With that connection between the notion of a reason that motivates and the notion

of a rationalising explanation in the background, we can now get a sharper grip on what the Motivating Reasons View amounts to. As has already been intimated, the Motivating Reasons View involves a commitment to the following two theses:

The Motivating Reasons Postulate. The notion of a reason which motivates the agent to ϕ is to be identified with the notion of a *motivating reason*. The notion of a motivating reason is the notion of an entity which has the following properties: (i) it plays a rationalising role, where its playing that role is identical to its playing the sort of motivational role at issue; (ii) it is a reason: we have a concept of a reason which applies to it; (iii) the rationalising explanation to which S's ϕ -ing is subject just in case S ϕ s motivated by the reason that p obtains across both the good and bad cases, so that the items which are motivating reasons can exist, counting as reasons, with the rationalising role being played, even if p is not a normative reason for S to ϕ ; (iv) motivating reasons are reasons in the sense of being items that play the relevant sort of motivating role.

The Primacy of Motivating Reasons. ϕ -ing in response to the normative reason that p reduces to having the motivating reason that p, understood in the way described by the Motivating Reasons Postulate, in addition to further factors present in the good case which are individually necessary and, together with having the motivating reason that p, jointly sufficient for ϕ -ing in response to the normative reason that p.

The Motivating Reasons Postulate identifies our ordinary notion of a reason that motivates with the special notion of a motivating reason. It then defines that special notion for us by appeal to four features.

First, the notion of a motivating reason is the notion of a reason which plays the rationalising role already described: it is the notion of an item which is, or is in some sense a part of, the *explanans* of the rationalising explanation to which S's ϕ -ing is subject when they ϕ for a reason. It has already been agreed that that is a correct thing to say about the items which we think of as reasons which motivate.

Second, motivating reasons really do count as reasons. That is to say: for any item we think of as a reason which motivates the agent to ϕ , there is a concept of a reason which is applicable to it. This, I think, is surely a trivial claim about the items which we think of as reasons which motivate.

Third, the rationalising explanation in terms of which we individuate what it is for S to ϕ motivated by the reason that p is an explanation that holds in both good and bad cases: it requires merely that the agent takes p to be a normative reason and is blamelessly responding to it accordingly, not that p actually is a normative reason to which S is responding. It follows that the items which are the reasons that motivate one to ϕ are present in the bad case and motivating one to ϕ in the bad case too: one ϕ s for the reason that p in both the good case and bad.

In the last chapter I noted that there is a way of providing a rationalising explanation which is common to good and bad cases. It is to use a psychologistic rationalising explanatory statement: ‘S ϕ ed/is ϕ -ing because S believes that p’. Whatever rationalising explanation is provided by the assertion of that sentence on a particular occasion, it is one which holds whether the subject is responding to a normative reason, or else merely purporting but blamelessly failing to do so. There is no reason to believe that there is more than one good case/bad case neutral rationalising explanation applicable to the subject’s ϕ -ing on any given occasion, and the proponent of the Motivating Reasons Postulate says that rationalising explanations associated with (1)-(3) are good case/bad case neutral. Thus, the proponent of the Postulate is committed to saying that the rationalising explanation reported by a relevant instance of (1)-(3) in terms of which they’ve individuated what it is to ϕ motivated by the reason that p, is identical to the rationalising explanation reported by a corresponding psychologistic rationalising sentence. In so far as an instance of ‘S is ϕ -ing for the reason that p’ reports a rationalising explanation, it reports the explanation reported by the corresponding instance of ‘S is ϕ -ing because they believe that p’. The proponent of the Motivating Reasons Postulate adds to this the distinctive claim that such rationalising explanations appeal to items which count as reasons – the reasons which motivate the agent to ϕ .

It follows from feature (iii) that motivating reasons needn’t be, and perhaps never are, normative reasons, for they are present in the bad case in which the agent doesn’t count as ϕ -ing in response to a normative reason. It also follows, given feature (ii), that motivating reasons are not reasons in the same sense in which normative reasons are. For if such items are present and count as reasons in whatever sense they do in bad cases, where the agent doesn’t count as ϕ -ing for a normative reason, then the sense in which the relevant item is a reason cannot be that it is a normative reason. It looks like the only option to which to appeal in order to account for the item’s status as a reason, then, is its standing in the motivating relation to the agent’s ϕ -ing. Since that is identified by (i) with its playing a certain rationalising role, it follows that motivating reasons are reasons in a distinct sense from the sense in which normative reasons are reasons, and that they count as reasons by dint of the rationalising-explanatory role they play. That’s just feature (iv). What this brings out is that if motivating reasons have features (i), (ii) and (iii), then it follows that they have feature (iv).

The Primacy of Motivating Reasons presupposes the Motivating Reasons Postulate and then goes on to assign a certain theoretical role to motivating reasons, so understood: we’re to appeal to them as part of a Reductive account of the good case. To offer a Reductive account of the good case is to say that we should think of the phenomenon of responding to a normative reason as partly reducing to the phenomenon of ϕ -ing in a way that is subject to the sort of good case/bad case neutral rationalising explanation reported using the phrase ‘S ϕ ed/is ϕ -ing because S believes that p’. We then add to that phenomenon the

factors which need to be added to it in order to ensure that S counts as ϕ -ing in response to a genuine reason. For example, we add the truth of p and that p favours S's ϕ -ing, as well as perhaps the fact that S knows that p, and so on.

The Primacy of Motivating Reasons says that we should run this Reductionist story about the good case, but adds that the story needs to be run in terms of the notion of a motivating reason, as understood by the Motivating Reasons Postulate. That's because it is those sorts of items which either are, or are a part of, the sorts of good case/bad case neutral rationalising explanations in terms of which the good case is being understood. So the proponent of Primacy says that ϕ -ing in response to a normative reason is a matter of having a motivating reason, so that one's ϕ -ing is subject to a neutral explanation of why one ϕ s, plus, for example, the truth and normative status of one's apparent normative reason as well, perhaps, as one's knowledge of the normative reason.

There is a point worth bringing out about Reductionist models of ϕ -ing in response to normative reasons of the sort embodied by the Primacy of Motivating Reasons thesis. Let's suppose that there is more than one type of rationalising explanation: more than one sort of way of explaining why an agent ϕ ed or is ϕ -ing that appeals to what can be said in favour of S's ϕ -ing from S's own point of view. The only reason for thinking so would be that responding to a normative reason involves a distinctive sort of rationalising explanation, one which is not good case/bad case neutral. But the proponent of a Reductionist model of the good case, as is the Primacy thesis being considered here, says that the only sort of rationalising explanation involved in the good case is a good case/bad case neutral explanation – one which involves the presence of Motivating Reasons, according to the Primacy thesis. So any Reductionist about the good case, including the proponent of the Primacy of Motivating Reasons, is committed to thinking that there is only one sort of rationalising explanation: the good case/bad case neutral one.

Moreover, we can add to this that if one is already convinced that there is only one sort of rationalising explanation: a good case/bad case neutral one, then one will also have to be Reductionist about the good case. For responding to a genuine reason constitutively involves one's ϕ -ing being subject to a rationalising explanation, and if that rationalising explanation must be of a neutral type, then we need to add factors also present in the good case in order to get us to the phenomenon of responding to a genuine reason. If one is operating with the Motivating Reasons Postulate, then the form of Reductionism will be the form encapsulated by the Primacy of Motivating Reasons.

So the correct thing to say is this. The claim that there is only one sort of rationalising explanation, the good case/bad case neutral one, is logically equivalent to a Reductionist account of the good case. The Primacy of Motivating Reasons is a Reductionist model of the good case. So if one adopts it, then one is committed to a monism about rationalising explanation. Moreover, if one commits oneself to a monism about rationalising explanation with an additional commitment to the Motivating Reasons Postulate in the

background, then one will thereby be committed to the Primacy of Motivating Reasons: the version of Reductionism subscribed to by the proponent of the Motivating Reasons View.

A comment on the logical relationship between the two theses which compose the Motivating Reasons View is in order. The Primacy of Motivating Reasons entails the Motivating Reasons Postulate. It is written in to the former that the latter is true, after all. However, it is possible to subscribe to the Motivating Reasons Postulate without subscribing to the Primacy of Motivating Reasons. That would involve accepting that being motivated by a reason is not simply a matter of responding to a normative reason but of there being a special sort of reason – a motivating reason – playing a motivating role for one, whilst denying that we should give an account of the good case in terms of what's common to both good cases and bad, so that there is only one fundamental sort of rationalising explanation, which is neutral on good cases and bad. At times, this seems to be the position which Hornsby wishes to occupy, for example:

It is certainly easy to put pressure on the idea of subjective reasons. But the pressure has to be resisted. Consider an example that Bernard Williams introduced: a man made a mixture of petrol and tonic because he wanted to drink gin and tonic and he believed that the petrol was gin. We may well think that there was no reason for the man to make the mixture. This is surely right if it means that there was no objective [that is: normative] reason for him to do so. But that does not mean that the man had no reason. (Hornsby, 2007a:90-91)

Here Hornsby seems to want to defend the thought that in bad cases in which it merely seems to the subject as if they have a normative reason to ϕ , there is something worth calling a reason on the scene nevertheless, which is a part of the rationalising explanation of their ϕ -ing. So Hornsby can be read as endorsing the existence of motivating reasons, as defined by the Motivating Reasons Postulate. But she would not want to endorse the Primacy of Motivating Reasons, for she thinks that in the good case, the rationalising explanation to which S's ϕ -ing is subject requires that S knows that p and hence isn't a type of explanation present in the bad case.

Hornsby's position is, however, equivocal, for sometimes she opts for factive readings of (1)-(3). So what should be said is that even though the Motivating Reasons Postulate doesn't entail the Primacy of Motivating Reasons, it at least isn't clear that there has been anyone wishing to occupy that position.⁴ In any case, it will be made clear later in the chapter that although the two claims which compose the Motivating Reasons View are not equivalent, they typically come together. I shall be suggesting that we reject both in the next chapter, nevertheless.

⁴Having said that, it's clear that Marcus (2012) accepts the Motivating Reasons Postulate whilst rejecting the Primacy of Motivating Reasons.

2.2 The Dialectic of Psychologism

I've introduced the Motivating Reasons View, but does anyone actually subscribe to it? If not, then the next chapter, in which I reject the View, might be said to lack significant philosophical interest. In this section I lay out a well known dialectic which arises in this area which I label *the Dialectic of Psychologism*. In particular, I lay out three major positions one might wish to occupy with respect to that dialectic: the Davidsonian View, Dancy's View and the Neo-Davidsonian View, each of which was briefly explored in the introduction. I aim to show that the Dialectic of Psychologism is permeated with the assumption that the Motivating Reasons View is correct. I aim to show this by demonstrating that the three views mentioned are actually distinct versions of the Motivating Reasons View.⁵ That suffices to prove that my target is real, whilst also serving to bring onto the scene a debate which I aim to resolve later. I describe the theory associated with Davidson in §2.2.1, the theory associated with Dancy in §2.2.2, and the Neo-Davidsonian Theory in §2.2.3.⁶

2.2.1 Davidson's View

This section provides a statement of Davidson's position, *as it is standardly understood*, and aims to show that, so understood, it is an instance of the Motivating Reasons View.⁷

Davidson's famous paper starts with the following question, which begins a passage I've already cited:

What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action. (Davidson, 1963:3)

The question amounts to this: what is the relation that holds between the agent's action and the *explanans* of the rationalising explanation of the agent's action? Davidson's answer is that the relation is causal:

In this paper I want to defend the ancient – and commonsense – position that rationalization is a species of causal explanation. (*Ibid.*)

⁵This raises the issue of what becomes of the Dialectic of Psychologism once the Motivating Reasons View is rejected, and I aim to explore that question at the end of Chapter Three and throughout Chapter Four. For now, all I want to do is lay-out the dialectic in question and demonstrate that each of the three major parties to it subscribe to the Motivating Reason View.

⁶It should be noted that these three options aren't exhaustive. Sandis (2011, 2013) presents us with a further option, for example, which is similar but importantly different from the third. I don't say much about that option in this essay, but it seems to be motivated partly by the truth of the Motivating Reasons Postulate, which I aim to refute in the next chapter.

⁷For the record, I think Davidson's actual view is not quite the view standardly ascribed to him. I ignore that important point in what follows.

Davidson thinks that the causal relation at issue links the agent's action with a set of psychological states of the agent which constitute the agent's seeing their action as something which is supported by what normative reasons there are. These psychological states are in turn identified with pairs of pro-attitudes and beliefs, which Davidson calls the agent's 'primary reasons', to the effect that there is some feature of the action by dint of which the performance of it promotes a goal to which the pro-attitude is directed:

Whenever someone does something for a reason, therefore, he can be characterized as (a) having some sort of pro attitude towards actions of a certain kind and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind. (Davidson, 1963:3)

Generalising to judgement sensitive phenomena across the board, the picture Davidson offers us of rationalising explanation is this. When the agent's ϕ -ing is subject to a rationalising explanation – as it is whenever they purport to ϕ in response to a normative reason – the rationalising explanation always takes on the form of an explanation which takes as *explanans* the concatenation of a state of believing that p and a pro-attitude state which together constitute the subject taking p to be a normative reason in favour of their ϕ -ing. Those states together cause the agent's ϕ -ing, and it is that relation of causation, linking the agent's ϕ -ing with their belief/pro-attitude pair, which underwrites the truth of a rationalising explanation of the agent's ϕ -ing.

Restricting my attention to the cognitive side of things, as I have been doing already, Davidson's view is that rationalising explanations are always to be given in terms of the agent's state of believing that p – where p is the item the agent takes to be a reason in favour of their ϕ -ing. For Davidson, then, the sense in which rationalising explanations appeal to what can be said in favour of ϕ -ing from the agent's own point of view is that they appeal to the psychological states which constitute, or at least partly constitute, the agent's taking the relevant factor to be a reason in favour of their ϕ -ing.

On the standard reading of Davidson, the reason which motivates the subject to ϕ is to be identified with the rationaliser of the subject's ϕ -ing. Davidson's talk of primary *reasons* is supposed to indicate that the beliefs which constitute the agent's rationalisers are the reasons which motivate the agent to ϕ . So, what goes for rationalisers also goes for the reasons that motivate the subject to ϕ , and *vice-versa*. Hence the reason which motivates the agent to ϕ are, for Davidson, to be identified with states of believing that p which cause the agent's ϕ -ing.

So far I've merely provided an exegesis of Davidson's position, as standardly understood, or at least the salient elements of it. What's important to note for my purposes, however, is that Davidson subscribes to the Motivating Reasons Postulate and to the Primacy of Motivating Reasons.

Take the former first. As I've already noted, it's quite natural to think that (i) is true of reasons which motivate, and I see no reason to think that Davidson, on the standard reading, would deny that the reasons which motivate exemplify such a feature. (ii) is trivially true of such items. (iv) follows from the conjunction of (i), (ii), and (iii), so now the question is: does Davidson think that (iii) holds true of the reasons which motivate?

Davidson identifies the reasons which motivate the agent to ϕ with beliefs the agent has directed towards the items which constitute what seem, from the agent's own point of view, to be reasons in favour of their ϕ -ing. Since such states of mind can be present and doing their causal work across the good case and bad, it follows that Davidson is committed to thinking that reasons which motivate can be present and doing the rationalising-explanatory work constitutive of the motivating role they play across good cases and bad. Thus, Davidson is committed to ascribing feature (iii) to the reasons which motivate too.

Now take the Primacy of Motivating Reasons. It's clear that Davidson thinks that rationalising explanation only ever takes the form of his good case/bad case neutral explanations which appeal to the relevant causally active states of mind. As I noted earlier, this commits one to a Reductive account of the phenomenon of responding to normative reasons anyway, and the shape of the Reductionism will have to be that specified by the Primacy of Motivating Reasons, given Davidson's prior commitment to the Motivating Reasons Postulate. In conclusion, then, Davidson, as standardly read, is indeed committed to the Motivating Reasons View.

2.2.2 Dancy's View

This section outlines the view Dancy defends primarily in his book *Practical Reality*.⁸ Dancy's View takes as its starting point the thought that Davidson is quite wrong to think that the reason which motivates the subject to ϕ is to be identified with the subject's belief that p. In general, Dancy rejects the thought that we should think of the reasons which motivate the subject to ϕ as being psychological states at all, whether states of believing that p, states of knowing that p, states of desiring to ψ , or whatever.⁹

But Dancy's rejection of the psychologistic tendency at the heart of the Davidsonian tradition goes even deeper. That's because we can, it seems, distinguish between the state of believing that p from the fact that the subject is in the state of believing that p. More generally: we can distinguish between states and facts which record the presence of those

⁸He also defends the view in Dancy (2003, 2004b, 2008, 2011). It should be noted, however, that Dancy (2014) modifies the original position substantially, so that it is more in-line with the view of Sandis (2011, 2013). I do not focus on Dancy's new position in his essay although, as I have already said, I think it is undermined by dint of the failure of the Motivating Reasons Postulate. In this section, I focus on the position defended in *Practical Reality* and I label it, somewhat misleadingly given what has just been said, 'Dancy's View'.

⁹Notice here that what Dancy is rejecting is not the claim that the reasons which motivate are narrowly supervening psychological states, but the more general thought that they are psychological states of some kind.

states.¹⁰ Thus, it is possible to deny that the reasons which motivate are psychological states whilst still suggesting that they are all psychological facts that record the existence of such states. Let us reserve the label ‘Psychologism’ for the view that the reasons which motivate the agent to ϕ are to be identified with psychological items of the subject, whether psychological states or facts.¹¹ Suppose I dial the number of my uncle for the reason that it’s his birthday. On the state version of Psychologism, the reason for which I dial is my believing that it is my uncle’s birthday. On the fact version of Psychologism, the reason for which I dial is that I believe that it is my uncle’s birthday. Dancy wishes to reject the fact version of Psychologism too: he wishes to reject Psychologism across the board.¹²

On Dancy’s alternative, we should think of reasons which motivate the agent to ϕ as what the agent takes to be the normative reason in favour of their ϕ -ing: that p . So the reason for which I dial my uncle’s number is not my believing, nor the fact that I believe, but rather what I believe: that it is my uncle’s birthday. Similarly, when I believe that Labour will win the by-election for the reason that the exit-poll predicts so, my reason is what’s designated by the that-clause: it is that the exit poll predicts so, not my believing it or the fact that I believe it. As Dancy says:

...I argue that we should not accept any form of Psychologism anyway. We should attempt to understand the reasons that motivate us as features of the situation rather than as features of ourselves. (Dancy, 2000:20)

I’ll call the claim that reasons which motivate the subject are not psychological items, either states or facts, but rather what the agent believes *Anti-Psychologism*. Dancy is the foremost proponent of Anti-Psychologism.

It should be flagged up that the Anti-Psychologist needn’t deny that sometimes the reason which motivates one to ϕ will be that one believes that p . They will just say that it gets to be so in so far as that one believes that p is the content of a (second order) belief that constitutes the presence to mind of what one takes to be a normative reason – that one believes that p . In such cases, the proponent of Psychologism will say that the reason

¹⁰ Although the distinction between states and facts is contentious, it is *prima facie* plausible, and I will go along with it here. States are the sorts of things which objects and organisms can be in, and, on the face of it, they occupy certain stretches of time over which they, like the objects and organisms they are states of, persist. Facts, on the other hand, are not the sorts of things which objects and organisms can be in, but rather are the sorts of things which involve or are about objects and organisms, and it is not obvious that facts are the sorts of things which occupy stretches of time at all, let alone persist through the stretches of time that they occupy. Thus, it seems plausible, for example, that we should distinguish the state of being red the post-box is in from the fact that the post box is red. And we should distinguish the state of believing that p a subject who ϕ s for the reason that p is in from the fact that the subject believes that p . It might be that those facts obtain if, and only if the relevant state is one which the corresponding object or subject is in. But the two sorts of entities should be distinguished nevertheless.

¹¹ Dancy reserves that label just for the state version of the view and reserves the label ‘the New Theory’ for the fact version.

¹² See Dancy (2000:Ch.6) for his rejection of the fact version of Psychologism.

which motivates one is one's second order state of believing, or else a fact which records the presence of such a second order state.¹³

For example, one might find oneself believing, quite irrationally, that it's likely that the aeroplane one is going to fly in is going to crash. One recognises that the belief is quite irrational, but one cannot shake it. Thinking about flying in the plane causes one to feel great panic and fear and whenever one thinks about it one finds oneself avowing, against one's better judgement, that the plane will crash. One decides to go and see a hypnotherapist in order to rid oneself of the irrational belief. The reason for which one visits the hypnotherapist is that one believes that the plane is likely to crash. That fact about oneself is present to mind in the way required for performing one's action on the basis of it because one has a (second order) belief that one believes that the plane is going to crash. The proponent of Anti-Psychologism can perfectly well accept the coherence of the case: they will identify the reason which motivates one with what one believes – that one believes that the plane is going to crash, not that the plane is going to crash. The Psychologist by contrast will say that the reason is one's second order belief that one believes the plane will crash. Anti-Psychologism asserts only that the reasons which motivates one are what one believes when one ϕ s for a reason, they needn't deny that one's reasons can be considerations about what one believes.¹⁴

So Dancy resists the Psychologistic tendency of Davidson, and indeed of the fact version of Psychologism too. The reasons which motivate are what the agent believes about the situation – the factors which the agent takes to be normative reasons, so that in the good case their reason just is what makes it so that they ought to ϕ . But, and importantly, Dancy agrees with Davidson that the reasons which motivate the agent to ϕ are identical to the rationalisers of the agent's ϕ -ing:

In this book I have argued against certain prevalent views about motivating reasons, the reasons for which we act. Such reasons are the things we appeal to when we try to explain (or at least to explain in a certain way) why someone did what he did. So a theory of such reasons contains, or is, a theory about the nature of a certain sort of explanation of action – and vice-versa. (Dancy, 2000:159)

¹³It is for this reason that when Turri (2009) contrasts what he calls 'Factualism', which is the Anti-Psychologistic view that the reasons which motivate are facts, and 'Abstractionism', which is the Anti-Psychologistic view that the reasons which motivate are the propositional contents of mental states thought of as distinct from facts, with what he calls 'Dualism', which is the view that reasons which motivate are sometimes mental items and sometimes worldly items, he is setting up a pair of straw men. Perhaps, however, all Turri needs is to be more careful with his formulation of the Anti-Psychologistic views he describes and how they contrast with his Dualism.

¹⁴For similar examples see Hyman (1999:444), Dancy (2000:125), and Alvarez (2010:48). The point here mirrors a point made in the last chapter, that normative reasons can sometimes be facts about the agent's own psychology, even though they are typically facts about the world. This is no accident, given that what motivates proponents of Anti-Psychologism is the thought that the reasons that motivate can sometimes just be the agent's normative reasons.

So the rationalisers of the agent's ϕ -ing are always to be identified with what the agent believes about the situation. When I dial my uncle's number, the reason that motivates me is that it is his birthday. That is also what explains why I do it, for Dancy. That goes for both the good case and the bad case: all rationalising explanations are explanations which appeal to the reason for which the agent ϕ s – p – as their *explanantia*. So here we have another disagreement with Davidson: the *explanantia* of rationalising explanations are always what the agent takes to be a normative reason, and are not psychological states which constitute their taking the relevant consideration to be a normative reason, nor corresponding psychological facts. That further disagreement is generated partly by an agreement with Davidson that the reasons which motivate are rationalisers and partly by the original disagreement with Davidson over whether reasons which motivate are psychological items.

Things start to get quite clearly controversial, with respect to Dancy's View, once attention is drawn to the fact that Dancy thinks that one can ϕ motivated by a reason, so that (1)-(3) are true of one, even in the bad case:

If I say... "the reason for which he was running was that the train was leaving", I can perfectly well continue by saying "in fact, however, he was quite wrong about that". (Dancy, 2000:346)

So Dancy's thought is that sentences (1)-(3), which report rationalising explanations that appeal to an item to which a concept of a reason applies, are true in both the good and bad cases, and across both sorts of case the item which is one's reason is held fixed: it is what one believes – p. That item just is the *explanans* of the rationalising explanation that holds of one across each case. But in bad cases it can be false.

It follows from this that rationalising explanations needn't have true *explanantia* – they constitute a counterexample to the claim that reasons why are always facts or truths. Thus, the rationalising explanations offered by (1)-(3) are explanations which can be true, even when their *explanantia*, namely instances of p, are false. And the same goes for every rationalising explanation, provided by any type of sentence capable of supplying such an explanation. This is consistent, of course, with the thought that there are ways of reporting rationalising explanations which do require the truth of p. Non-Psychologistic rationalising statements that use the 'because' idiom and its cognates fit that description. But for Dancy, in so far as such statements report rationalising explanations, they do not require the truth of p. They might require that p be true, but that is not a function of the fact that they provide rationalising explanations.¹⁵

Since causal explanations require the truth of their *explanantia*, it follows, in turn, that rationalising explanations are not causal explanations, *contra* Davidson. It is not so that

¹⁵See Dancy (2000:Ch.6.3). We will be looking in more detail at what Dancy has to say about the 'because p' idiom in the next chapter.

when one reports a rationalising explanation of why S ϕ s the state of affairs which one has in mind is a state of affairs involving a causal link between the agent's ϕ -ing and whatever is doing the explaining. Instead, Dancy thinks that there is a *sui generis* ' ϕ -ing for a reason' relation that holds between the agent's ϕ -ing and the reason for which the agent ϕ s. This relation is non-causal, it links the agent's ϕ -ing with their reason – what they take to be a normative reason, and it doesn't require the truth or normative status of the agent's reason. It is that state of affairs we have in mind when we think of a rationalising explanations as applying to the subject's ϕ -ing, as we do when we apply (1)-(3) to the subject and when we apply psychologicistic and non-psychologicistic rationalising statements to the subject too.¹⁶

I've offered a brief characterisation of Dancy's position. But is it a version of the Motivating Reasons View? I think it is, as I now want to argue.

Take the Motivating Reasons Postulate first. Dancy certainly uses the phrase 'motivating reason' and he applies that phrase to the ordinary phenomenon of the reason that motivates one to ϕ . But this doesn't add up to a commitment to the Motivating Reasons Postulate unless he thinks of motivating reasons as having features (i)-(iv). I think it's clear from quote from p.159, given above, that he ascribes to motivating reasons feature (i): the motivating role played by motivating reasons just is a rationalising role. Feature (ii), moreover, trivially holds of motivating reasons. We've also seen that feature (iii), in conjunction with (i) and (ii), gets us feature (iv).

So now the question is: does Dancy ascribe feature (iii) to the reasons which motivate? That is: does he think of the reasons that motivate one to ϕ as items which constitute part of a rationalising explanation that is true both in the good case and in the bad? The answer to this, of course, is yes. For Dancy, rationalising explanation always consists in an explanation that's grounded in his *sui generis*, non-causal, non-factive 'for the reason that' relation. That relation is present whenever the agent is purporting to ϕ in response to a normative reason – even if they blamelessly fail. And it links the agent's reasons – what they believe – with their ϕ -ing, so that the agent's reasons are rationalisers.

Now let's see whether Dancy subscribes to the Primacy of Motivating Reasons. Well, Dancy acknowledges only one sort of rationalising explanation: that which is grounded in the holding of his *sui generis*, non-causal, non-factive 'for the reason that' relation. That relation obtains across both the good case and the bad case and links the same *relata* in them. But as I've already noted, this suffices for one to be committed to some kind of Reductionism about the good case. Given his background commitment to the Motivating Reasons Postulate, Dancy will have to subscribe to the form of Reductionism specified by the Primacy of Motivating Reasons. And indeed that sort of account does seem to be the account Dancy goes for. He says that in the good case, on top of the sort of neutral rationalising explanation he envisages holding, the reason for which one ϕ s is identical to a reason in favour of one's ϕ -ing, and that is all he has to say about what is involved in

¹⁶See 2000:Ch.8.

the good case. That certainly makes it seem as if what he is up to is accounting for the good case in terms of having a motivating reason – which consists in the holding of his primitive, non-causal, non-factive relation – plus, over and about that, the truth of p.

It might be objected to this reading of Dancy that it doesn't square well with what he says here:

If we do speak in this way, of motivating and normative reasons, this should not be taken to suggest that there are two sorts of reason, the sort that motivate and the sort that are good. There are not. There are just two questions that we use a single notion of a reason to answer. When I call a reason 'motivating', all that I am doing is signalling that our attention is on matters of motivation, for the moment. When I call it 'normative', again all that I am doing is stressing that we are currently thinking about whether it is a good reason, one that favours acting in the way proposed. (Dancy, 2000:2-3)

What Dancy seems to say here is that the distinction he draws is not intended to be a distinction between two kinds of reasons: normative and motivating. I have attributed to him the thought that normative and motivating reasons are different in kind in the sense that the concept of a reason associated with normative reasons is distinct from the concept of a reason associated with motivating reasons. I haven't attributed to him the thought that they are different in kind in the sense that they belong to different ontological categories – that's no part of the Motivating Reasons View. So if such a metaphysical claim is what Dancy has in mind when we says that they are the same in kind, then that's no problem for my interpretation.

However, Dancy is quite clear that he thinks that there is only one concept of a reason in this area, which we use when posing two different questions. So isn't my reading of him under threat?

I don't think it is because, first of all, as I have argued, the details of Dancy's theory commit him to thinking that normative and motivating reasons are reasons in different senses of the word. Even if he wishes to deny that claim, his theory commits him to it nevertheless. It commits him to it because reasons that motivate are present in the bad case, and in the bad case those items really are reasons. But if that's so, then the sense in which they're reasons is not that they are normative – for bad cases include cases in which there is no normative reason present.

In any case, it's not at all clear why the fact that there is a notion of a reason which figures in a normative question and a notion which figures in a motivational question should imply that those notions are identical. Of course, that fact is *consistent* with the thought that there is really only one notion here – which is my preferred view. But it doesn't imply it. Thus, it's not clear why pointing to that fact should give us any reason to think that

there is only one notion of a reason here, as Dancy seems to think. I conclude, then, that Dancy really does subscribe to the Motivating Reasons View.

2.2.3 The Neo-Davidsonian View

I think it is fair to say that Dancy's resistance to Psychologism has been accepted by many philosophers working in this area. It is also fair to say that few philosophers, including those who accept Dancy's Anti-Psychologism, have wished to follow Dancy in saying that rationalising explanations are non-factive and hence non-causal. It is quite intuitive that explanations why in general are factive, and that rationalising explanations are explanations why. But how are we to combine the factivity of explanation with the claim that the reason for which the agent ϕ s are what the agent takes to be normative reasons?

Well, one way to do it is to deny that (1)-(3) are true in the bad case, and to counteract whatever supposed reasons there are for thinking that that's so, so that (1)-(3) pick out the phenomenon of responding to a normative reason. That way, one can say that the rationalising explanation reported by (1)-(3) take the reason for which the agent ϕ s as *explanans*, but the agent's reason is a fact in such a situation, so we don't have a violation of Factivity. In the bad case there is a different sort of rationalising explanation available, one which appeals to the fact that the agent believes that p. So there are two forms rationalising explanation. With respect to one of them, the agent's reason is the rationaliser and it is non-psychological. With respect to the other, there is nothing that counts as a reason which forms part of the rationaliser, and the rationaliser is the fact that the agent believes that p. That is the strategy which, ultimately, I want to defend, but it involves rejecting the Motivating Reasons View, as we shall see in more detail in the next chapter.

Here I want to focus on an alternative strategy for marrying Anti-Psychologism with Factivity. On this strategy, we are to accept that (1)-(3) are true in the bad case as well as the good. Hence we are to follow Dancy and Davidson in thinking of the rationalising explanations reported by instances of (1)-(3) as good case/bad case neutral. But we should follow Davidson in saying that the *explanantia* of rationalising explanations are the beliefs directed towards the reason held by the agent, or else a fact which records the presence of such a state. We can combine that with Dancy's plausible thought that the reasons for which one ϕ s are what one takes to be normative reasons by rejecting the thought, shared by both Davidson and Dancy, *that the reason for which one ϕ s is identical to the rationaliser of one's ϕ -ing*. The reason for which the agent ϕ s still counts as *part of* the agent's rationaliser. But it isn't identical to it, as on the Davidsonian and Dancyan Views.

This is the strategy adopted by Smith (2003), Davis (2003, 2005), Olson and Svensson (2005), Setiya (2011), and Hieronymi (2011). I label that view the *Neo-Davidsonian View*. It is an attempt at synthesising Dancy's View with Davidson's View, at least in certain respects. Here is a representative statement of the basic idea from Davis:

We need to distinguish the claim that actions can be explained by reference to reasons from the claim that the reasons are what explain the actions. The former is true, the latter false. The statement that my reason for saving was that my son will need money for college does explain why I saved. But it does not follow, and it is not true, that my reason explains my action. For my reason was that my son will need money for college. That something will be true in the future cannot explain the fact that I did something in the past. Moreover, my action would have the same explanation even if I were wrong in thinking that my son will be going to college. (Davis, 2003:455)

What is the *explanans* of the rationalising explanation to which one's ϕ -ing is subject when one ϕ s for a reason, for Davis? He argues that ϕ -ing for the reason that p – the sort of phenomenon picked out by (1)-(3) – consists in ϕ -ing in a way caused by the psychological states which constitute one's taking p to be a reason in favour of one's ϕ -ing. The *explanans* of the rationalising explanation to which S's ϕ -ing is subject, for him, then, must be the psychological states in question, or at least facts which record the existence of them. But he's clear that he accepts that the reason for which the agent ϕ s is that p, and indeed that the reason for which the agent ϕ s is that p in both the good case and the bad – indeed, this is one of the reasons cited in the above passage for thinking that such reasons can't be rationalisers: when such reasons are false, the relevant *explanans* is not identical to them, and remains the same as it is in the good case. So for Davis the reason for which S ϕ s is not identical to the rationaliser of their ϕ -ing. ϕ -ing for the reason that p can consist in one's ϕ -ing being caused by elements of one's psychology, so that those elements are the rationalisers of one's ϕ -ing, even though those elements are not the reasons for which one ϕ s. The reason for which one ϕ s still figures in the rationalising explanation of one's ϕ -ing, it just does so by being what's believed by the agent, where it is the belief which constitutes the presence to mind of one's reason which is one's rationaliser, on Davis' picture.¹⁷

The position defended by Setiya (2011) is much the same. First, we can note Setiya's subscription to Anti-Psychologism:

The doctrine of 'psychologism', according to which the reasons for which we act are always states of, or facts about, our own psychology, is hopeless. (*Ibid.*:132)

Next, we can note that he allows that S can ϕ for the reason that p even in the bad case:

... the fact that I am doing ϕ on the ground that p is consistent with its being no [normative] reason at all for me to ϕ . (*Ibid.*)

¹⁷Sandis (2013) and Dancy (2014) think that Davis wishes to defend their own distinctive position which purports to deny that the reasons for which the subject ϕ s are psychological entities, that rationalisers are psychological entities, *and* that rationalisers are identical to the agent's reasons, thus preserving Factivity. I think they misread Davis, who clearly thinks that rationalisers are psychological items.

Finally, we can note that Setiya thinks, like Davis, that ϕ -ing for the reason that p consists in ϕ -ing in a way that's caused by the states of mind which constitute one's taking p to be a reason. Rationalisers, for Setiya, then, must be the states which cause. But they are not the reasons for which one ϕ s:

...[Dancy] seems to infer...that the failure of psychologism, so understood, casts doubt on causal-psychological accounts of acting for a reason. This is not the case. That agent's reasons are not psychological states and need not be facts about their psychology leaves open that, in order to act on the ground that p , one must have certain psychological states, that these states must play a specific 'efficient causal role', and that this is *what it is* to act for that reason. (Setiya, 2011:133)

We find a further proponent of Neo-Davidsonianism in Hieronymi:

We now have on stage three contenders for the title 'reasons for action': first, considerations that (in fact, truly) count in favour of acting, which Scanlon calls 'reasons in the standard, normative sense' and others sometimes call 'normative reasons'; second, considerations that someone *took* to count in favour of acting, on the basis of which he or she acted (considerations someone treated as reasons in the standard normative sense)...which are sometimes referred to as 'the agent's reasons' (and often as 'motivating reasons'); and, finally, considerations that explain an action, whatever these may be. Given the possibility of error, this last role seems to be played by psychological states of the agent...by keeping in mind that there are (at least) these three possible characters, we can minimize confusion. (Hieronymi, 2011:413)

Here Hieronymi suggests that we should distinguish three things the phrase 'reason for action' might stand for. First, there are normative reasons. Second, there are the reasons which motivate the agent to ϕ , which are given Dancy's Anti-Psychologistic treatment. Then there are the *explanans* of rationalising explanations – a certain sort of reason why S ϕ s. The latter are identified with psychological states by her, and the reason given for that is that the subject's reasons for ϕ -ing can be present in the bad case, but, given Factivity, they can't explain because they're false. Thus, the agent's reasons are not identical to the agent's rationalisers, for Hieronymi. She takes it that we can perfectly well distinguish between the three sorts of reasons she describes – there is no problem in doing so, and those who conflate the latter two sorts of reasons like Davidson and Dancy are simply confused. Much the same position is recommended by Olson and Svensson (2005).

Some Neo-Davidsonians, such as Smith, Davis, Setiya and Olson and Svensson, go along with Davidson in thinking that rationalising explanation is causal. Hieronymi does

not commit to that, and I do not build a commitment to Causalism into the Neo-Davidsonian position. Moreover, there isn't agreement amongst the Neo-Davidsonian camp about precisely what the *explanantia* of rationalising explanations are. Some, like Smith, Setiya, Davis, and Olson & Svensson, seem to be straight-up Davidsonians at the level of rationalising explanation. Others, like Hieronymi, wish to leave it open to some extent which psychological states constitute rationalisers. I do not build into the Neo-Davidsonian position any particular conception of rationalisers. But I do build it into the position that: (a) the rationalisers are not identical to the reasons for which the subject ϕ s; (b) the rationalising explanations are good case/bad case neutral; and (c) they appeal to states of mind of the subject which constitute the subject taking p , the reason for which they ϕ , to be a normative reason for them to ϕ , or else to facts which record the presence of such states.

Why are the Neo-Davidsonians subscribers to the Motivating Reasons View? Take the Motivating Reasons Postulate first. Several of the authors mentioned use the phrase 'motivating reasons', or else something similar, to refer to the reasons which motivate the agent to ϕ . But do they think of such things as exemplifying features (i)-(iv) of the Motivating Reasons Postulate? Well feature (i) is quite natural, as has already been noted, and none of the Neo-Davidsonians show any sign of recognising that it might be possible to deny it. Feature (ii) is trivial. Feature (iv) follows from the conjunction of (i), (ii) and (iii).

So now the question is: do the Neo-Davidsonians think of the reasons that motivate as having feature (iii)? I think the answer is yes. For on the Neo-Davidsonian View, the reasons which motivate the agent to ϕ are constituents of the *explanantia* of rationalising explanations, even though they are not identical to such *explanantia*. And the explanations in question are good case/bad case neutral: they are given in terms of the psychological states or facts which constitute the appearance of normative reasons to the subject, where those states of mind are present across both the good and bad cases. So the Neo-Davidsonians really are proponents of the Motivating Reasons Postulate.

Now let's take the Primacy of Motivating Reasons. The Neo-Davidsonians don't discuss what account to give of the good case very much, so far as I can tell, although many of them – Olson and Svensson excepted – are sympathetic to Dancy's claim that in the good case one's reason is identical to the reason which favours one's ϕ -ing. However, one thing is clear: all the Neo-Davidsonians are in agreement that there is only one sort of rationalising explanation – the sort of good case/bad case neutral rationalising explanations which Davidson envisages which appeal to certain states of mind. Thus, they must be Reductionists about what ϕ -ing in response to a normative reason consists in. Given their commitment to the Motivating Reasons Postulate, moreover, this Reductionism must take the form encapsulated by the Primacy of Motivating Reasons.

2.3 The Way Ahead

I've laid out the motivating reasons view and I've described the Dialectic of Psychologism, in particular the three main views one can take with respect to that dialectic: the Davidsonian View, Dancy's View, and the Neo-Davidsonian View. I've attempted to demonstrate that the Dialectic of Psychologism is permeated by the assumption that the Motivating Reasons View is correct, so that the three views mentioned are cast in terms which either presuppose or otherwise commit their proponents to the Motivating Reasons Postulate and to the Primacy of Motivating Reasons.

Although I have focused on reading the Motivating Reasons View into the authors who are party to the Dialectic of Psychologism, I hope the reader will recognise that the View is something of a mainstay of the contemporary analytic literature on the phenomenon of ϕ -ing for reasons, both in the case of believing for reasons and in the case of acting for reasons. A great many authors distinguish normative from motivating reasons, where the latter are identified with the ordinary phenomenon of the reason which motivates one to ϕ . And many of those authors think of motivating reasons as items which play an explanatory function but which are present whether or not the subject is ϕ -ing for a genuine normative reason. And, I think it's plausible that a great many writers only envisage one sort of rationalising explanation, thus committing them to a Reductionist model of the good case which typically takes the form of the Primacy of Motivating Reasons, given a prior commitment to the Motivating Reasons Postulate. Refuting the Motivating Reasons View, then, is of no small consequence. At the very least, it would require a great many philosophers to significantly recalibrate their views, but I will show that it will require more than just that.

At this point I want to say something about how I'm going to proceed. In the next chapter I'm going to attack the Motivating Reasons Postulate: I'm going to aim to show that the reasons which motivate the agent to ϕ are not items which exemplify features (i)-(iv) that are definitive of being a motivating reason in the special sense indicated. In particular, I'm going to argue that (1)-(3) report the success condition of ϕ -ing in response to a normative reason, and hence that the concept of a reason applied by someone who endorses one of them is distinct from the concept of a normative reason. Since there are no other sorts of sentences which are true in both the good cases and the bad which do apply a concept of a reason to anything which is supposed to play a rationalising role specified by the sentence in question, there is no reason to think that there are such things as motivating reasons in the special sense indicated.

I argue at the end of the next chapter, however, that although that result refutes the Motivating Reasons View by refuting the Motivating Reasons Postulate, that is certainly not the end of the debates to which Motivating Reasons Theorists have applied themselves. In particular, the Dialectic of Psychologism lives on even after the presupposition of the

Motivating Reasons View is removed. I reformulate that dialectic at the end of the next chapter in a way that's free of that false assumption and in Chapter Four I attempt to address that dialectic, settling the matter in favour of Anti-Psychologism and going some way toward settling it in favour of the Dancy-Davidson thesis that the reasons which motivate the agent are identical to the agent's rationalisers. As we will see, the reasons to reject the Motivating Reasons Postulate help to arrive at the correct resolution of the reformulated Psychologism Dialectic.

Moreover, although the Primacy of Motivating Reasons is refuted in the next chapter, given that it presupposes the Motivating Reasons Postulate, I aim to show that even with the Motivating Reasons View refuted, it remains a live issue whether to account for the good case in terms of what's common to the good and bad cases, plus other factors present in the good case over and above what's common – a position which we've seen is logically equivalent to the claim that there is only one sort of rationalising explanation. I address that issue, again free of the false assumption in favour of the Motivating Reasons View, in Chapter Five, where I argue against the Reductionist claim.

Chapter 3

The Myth of Motivating Reasons

Suppose I get a taxi to the airport for the reason that my flight to Belgium leaves soon. Would I still count as catching the taxi for that reason if I were blamelessly mistaken that my flight leaves soon? That is, is ϕ -ing for the reason that p a condition that obtains across both good and bad cases? The orthodox answer to this question is affirmative, indeed an affirmative answer to that question has been taken to be *obvious*, or at least quite easy to prove.¹ But once we've given an affirmative answer to the question, we're on our way to committing ourselves to the Motivating Reason Postulate. For if reasons which motivate us to ϕ are present, counting as reasons and doing their motivational work even in the bad case then there are reasons which motivate agents across good cases and bad but which don't count as reasons in the sense of being normative reasons.

This chapter argues that ϕ -ing for the reason that p is a matter of ϕ -ing in response to a normative reason, so that ϕ -ing for the reason that p is not a condition which obtains in the bad case. With that claim rejected, we're thereby in a position to reject the Motivating Reasons View by rejecting the Motivating Reasons Postulate. This will require us to re-assess the Psychologism Dialectic, as well as the debate about whether to be Reductionist about the good case, stripped of the usual presumption in favour of the existence of motivating reasons. Those tasks are begun here but are completed only in later chapters.

The style of argument to be developed here for the claim that ϕ -ing for the reason that p is true only in the good case is a familiar style of argument that comes in two stages. First, I establish that the thought that one can ϕ for a reason only in the good case should be our default view: a view which has the status of innocent until proven guilty. Second, I attempt to undermine arguments to the contrary.

I shall proceed as follows. In §3.1 I outline my own preferred interpretation of sentences (1)-(3) a consequence of which is that they are true only in the good case, and I try

¹See, for example, Smith (1987), Parfit (1997), Turri (2009), Miller (2008), and Setiya (2011) for the obviousness of the claim and Dancy (2000, 2003, 2011), Schroeder (2008), and Comesăna and McGrath (2014) for some attempts to prove the thesis with ease. Typically, the claim is simply presupposed without a commitment to it explicitly being flagged-up, perhaps because it is treated as obvious or else easily provable by the relevant philosopher.

to show that that reading of them should be our default view, thus shifting the burden of proof onto my opponent. In §3.2 I undermine an argument owed to Dancy (2000) which drives the thought that it is *obvious* that the contrary reading of (1)-(3) is true: the argument that it's simply intuitive that we can coherently conjoin a denial of p with each of (1)-(3). In §3.3 I pause to address a dialectical worry that arises out of consideration of that argument. In §3.4 I return to the task of refuting extant arguments for my opponent's position by attempting to refute a well-developed form of argument offered by Schroeder (2008) and separately by Comesăna and McGrath (2014) which appeals to connections between ϕ -ing for a reason and ϕ -ing rationally. In §3.5 I dispatch the simple argument that it is simply incredible to claim that the agent in the bad case ϕ s for no reason at all. In §3.6 I aim to undermine an argument for the claim that (1)-(3) can be true in the bad case which appeals the thought that we are always in a position to tell whether or not we are ϕ -ing for the reason that p . That concludes my defence of my preferred understanding of (1)-(3). In §3.7 I move on to assess what results the claim that (1)-(3) are true only in the good case has for the Motivating Reasons View and I argue that it refutes it by refuting the Motivating Reasons Postulate. In §3.8 I suggest that rejecting the Motivating Reasons Postulate doesn't preclude the Dialectic of Psychologism from getting under-way, and nor does it preclude the possibility of a Reductionist account of the good case. I offer cashings out of those two issues free of the usual assumption in favour of the Motivating Reasons View, so that I can pursue each of them afresh in future chapters. Finally, in §3.9 I address a kind of scepticism about the debate to be engaged in here: that it is a merely linguistic affair with little philosophical significance. If the reader is sympathetic to that sort of scepticism at the outset then I ask them to suspend it until the final section.

3.1 Being Motivated by a Reason as Responding to a Normative Reason

Consider again sentences (1)-(3):

- (1a) S ϕ ed/is ϕ -ing for the reason that p
- (1b) The reason for which S ϕ ed/is ϕ -ing is that p
- (2a) S ϕ ed/is ϕ -ing on the basis of p
- (2b) The basis on which S ϕ ed/is ϕ -ing is that p
- (3a) S ϕ ed/is ϕ -ing on the grounds that p
- (3b) The grounds on which S ϕ ed/is ϕ -ing is that p

These sentences record the fact that the agent has ϕ ed or is ϕ -ing motivated by the reason that p, where being motivated by a reason constitutively involves ϕ -ing in a way that is subject to a rationalising explanation. The rationalising explanation in question takes one's ϕ -ing as *explanandum* and takes, as its *explanans*, some condition with either is, or at least involves, the reason for which the agent has ϕ ed or is ϕ -ing specified by the relevant instance of (1)-(3).

On what I think is a perfectly intuitive reading of (1)-(3), they ascribe the property of being a normative reason to the item that motivates and then say of the subject that they are responding to that reason, so that the factor which is the agent's normative reason is, or is at least part of, the rationaliser of the agent's ϕ -ing. On this view, the sense in which the reasons that motivate are *reasons* is just that they are normative reasons. And on this view (1)-(3) are false in the bad case even though the agent who reflects on their situation in the bad case would mistakenly, though blamelessly, come to the conclusion that (1)-(3) are true of them.

If we go for this reading of (1)-(3), the phenomenon of a reason which motivates the agent to ϕ is nothing over and above the phenomenon of ϕ -ing in response to a normative reason, which was introduced in Chapter One. Thus, if we go for this reading, the introduction of the phenomenon of being motivated by a reason doesn't extend the debate in this area further than the thought that there is such a thing as responding to a normative reason. There would then be a debate about what that amounts to. I call the reading in question *the factive reading*.

To be clear, the factive reading of (1)-(3) is intended to be identified with the claim that (1)-(3) record the success condition of ϕ -ing in response to the normative reason that p. The factive reading entails (i) that the concept of a reason utilised by (1)-(3) is that of a normative reason, (ii) that (1)-(3) require the truth of p and its normative status, and (iii) that (1)-(3) are false in the bad case. Indeed, one of the major consequences of the factive reading is that the subject in the bad case does not ϕ for a reason at all. For to say that they do so is to say that they ϕ for a normative reason. Yet that is precisely not what happens in the bad case. According the contrary reading I wish to attack, (1)-(3) do not record the presence of the success condition, so that they can be true even in the bad case. I call that reading *the non-factive reading*.

The factive reading of (1)-(3) is preferred, for example, by Stout, who in the context of a discussion of a case in which a subject, Sally, is running for the reason that a bear is chasing her, focuses on the locution 'Sally's reason for running is that such-and-such'. He labels that claim (C) and has this to say about it:

The reasons given in formulations like (C) seem to be normative reasons that motivate. (Stout, 2009:55)

The factive reading is also preferred by Alvarez, who is quite clear on her commitment

to the claim that in the bad case the subject doesn't ϕ for a reason at all:

... it is right to say that an agent who acts motivated by a false belief does not act for a reason; rather he acts for a purpose and is motivated by an *apparent* reason. (Alvarez, 2010:146)

Others who endorse it include Raz (2011), Roessler (2014) and Hyman (1999, 2006, 2010).² Deniers of the factive reading must include all of those who subscribe to the Motivating Reasons Postulate, for the factive reading, as we'll see, is inconsistent with the Postulate.

Is the factive reading of (1)-(3) really so intuitive? Can anything be said to help elicit the intuition in favour of it? Well let's take an instance of the present tense, progressive form of (1a):

(1a*) Smith is walking to the quad for the reason that there is free coffee available there

One way of eliciting the intuition in favour of the factive reading of (1a*) is to compare it with the following instance of a non-psychologistic rationalising sentence:

(1_{bc}) Smith is walking to the quad because there is free coffee available there

(1_{bc}) clearly requires a factive reading: what comes after the rationalising 'because' contained in the sentence must be a fact which favours Smith's walking to the quad, and Smith needs to count as responding to that normative reason as such, if (1_{bc}) is to be true at all. That's just to say that non-psychologistic rationalising sentences report the condition of ϕ -ing in response to a normative reason.

But intuitively, (1a*) and (1_{bc}) say the same thing: they are just two different forms of words which function to express the very same thought. When Smith reports why he is walking to the quad, he could use either sentence, or use one and then use the other, and we would take him to be reporting the very same thing. More generally any sentence of the form: 'S has ϕ ed/is ϕ -ing for the reason that p' means the same as a sentence of the form: 'S has ϕ ed/is ϕ -ing because p', where the 'because' which appears in the latter sort of sentence is a rationalising 'because'. Thus, given that the factive reading holds of (1_{bc}), it must also hold of (1a*) too. The latter reports the success condition just as its 'because' counterpart does.

This result generates the following dialectical situation. Our default view in this area should be that the factive reading of (1)-(3) is accurate. The onus of proof is thus on those who wish to allow that (1)-(3) are true even in the bad case. If it can be shown, therefore, that there are no good reasons to believe that the factive reading should be rejected, then

²As we've seen, Hornsby is equivocal with respect to whether she thinks that the agent ϕ s for a reason only in the good case.

we are within our rights to conclude that the factive reading is true. During the course of this chapter I aim to disprove the four most compelling arguments in favour of a non-factive reading of (1)-(3), beginning with an argument given its clearest expression in Dancy (2000).³ Immediately after the discussion of Dancy, I will also seek to address a worry about the dialectical strategy being pursued here.

3.2 Dancy's Master Argument Rejected

Dancy has done much to promote the non-factive reading of (1)-(3). Here's what Dancy has to say in favour of his preferred reading:

I suggest that locutions such as:

His reason for doing it was that it would increase his pension

The ground on which he acted was that she had lied to him

are not factive. To test this we only need to consider whether it is possible without contradiction to continue by denying that things were as the agent took them to be. Consider the following sentences:

His reason for doing it was that it would increase his pension, but in fact he was quite wrong about that.

The ground on which he acted was that she had lied to him, though actually she had done nothing of the sort.

Neither of these sentences sounds self-contradictory to me. Not everyone's ears agree with me about this, I know. But there seems to be no reason why there should not be a way of revealing the light in which the agent saw things as a way of explaining why he did what he did, but without asserting that he was right to see things that way. I think that the two locutions above are ways of doing that. (*Ibid.*:132-133)

I think there are really two separate arguments contained in the above passage. The first argument is that intuitively, one can conjoin, with coherence, any of (1)-(3) with the denial of p. It follows that the truth of (1)-(3) doesn't require that p is a reason in favour of the subject ϕ -ing, so that (1)-(3) cannot record the success condition of responding to the normative reason that p. Much the same reasoning is defended by him elsewhere.⁴ I'll be spending most of my time undermining that argument, but before I do so I want to quickly dispatch the second argument I think is contained in the passage.

³I explore and dispatch a fifth and final argument in Chapter Five.

⁴See, for example, Dancy (2003, 2004b, 2011, 2014).

The second argument is expressed by the sentence beginning “But there seems to be no reason why there should. . .”. The thought seems to be that we can provide rationalising explanations which do not commit us to the truth of what the agent takes to be a normative reason, and that that provides support in some way for the non-factive reading of (1)-(3). It is of course true that we can provide the sorts of neutral rationalising explanations at issue: the sorts of explanations reported by psychologistic rationalising statements do just that. But I do not see how this could support the non-factive reading of (1)-(3). To achieve such support Dancy would have to appeal to the claim that (1)-(3) report the same rationalising explanation as psychologistic rationalising statements. For only with that claim in tow would it follow that the truth of (1)-(3) does not require p to be a reason in favour of S ’s ϕ -ing. But the proponent of the factive reading will precisely deny that (1)-(3) report neutral rationalising explanations. The second argument, then, should be dismissed.

I now wish to return to the first argument, according to which it’s intuitive that the denial of p can coherently be conjoined with any sentence of the (1)-(3) forms. The problem with Dancy’s argument is that he conflates (1)-(3) with the following, quite different sorts of sentences, each of which, we can safely assume, mean the same thing:

- (4) S ’s reason for ϕ -ing is that p
- (5) S ’s basis for ϕ -ing is that p
- (6) S ’s ground for ϕ -ing is that p

The conflation is evidenced in the passage cited, but I believe it to run through all of Dancy’s work on this matter. In the passage in question Dancy focuses on a pair of sentences. The first in the pair – “his reason for doing it is that it would increase his pension” – is an instance of (4). The second – “the ground on which he acted was that she had lied to him” – is an instance of (3b). Dancy takes it for granted that they have the same meaning. More generally, he takes it for granted that (1)-(3) have the same meaning as (4)-(6).

Let’s focus on (4) and take the following example:

- (4*) Smith’s reason for walking to the quad is that there is a free coffee available there.

What I’m charging Dancy with is committing himself to the following valid but unsound piece of reasoning: (4*) doesn’t require the truth of p , (4*) always means the same as (1a*), therefore (1a*) doesn’t require the truth of p . The argument for the claim that (4*) doesn’t require the truth of p is just the appeal to the intuition that we can conjoin (4*) with the denial of the claim that there is free coffee available in the quad without any incoherence. The claim that (4*) must mean the same as (1a*) is taken for granted. The

argument can be iterated for any of (4)-(6) and their (1)-(3) counterparts yielding the desired conclusion that (1)-(3) don't require the truth of p and hence that the factive reading is incorrect.

The piece of reasoning is unsound because it is not true that (4*) must mean the same as (1a*) and more generally it is not true that (4)-(6) must mean the same as (1)-(3). What I now want to argue is that (4)-(6) admit of two readings. On the one hand, they can be read so that they do mean the same as (1)-(3), in which case they record the presence of the success condition. On the other hand, they can be read so that they mean something quite different to (1)-(3), in which case a non-factive reading is true of them. Dancy's argument is undermined either way, as I shall show.

Sentences (4)-(6) tell us that the subject is ϕ -ing in response to something which is qualified by the phrase 'S's reason', a phrase which I take to mean the same as 'S's own reason'. (4*), for example, says of Smith that he is walking to the quad, and it reports to us that he's doing that in response to something that is qualified as *his own reason* for doing it – as *Smith's* reason for doing it. The sentences in question, I take it, report rationalising explanations which take as *explanantia* the item that is qualified as the subject's own reason, or else a condition which involves the item in question. Any interpretation of (4)-(6) must accommodate that fact, and also provide an account of what it amounts to to qualify the item in question as the subject's own reason.

On the factive reading of (4)-(6), the sense in which the relevant item is said to be a reason, basis or ground is that it is a normative reason: a fact which favours S's ϕ -ing. (4)-(6) are then construed as reporting the success condition of responding to a normative reason, just as (1)-(3) do on the factive reading. Thus, on the factive reading, (4*) says that Smith is walking to the quad in response to the reason that there is free coffee available there. The point of labelling the normative reason in question the subject's own is merely to emphasise that of all the normative reasons possessed by the subject who ϕ s, the subject has selected that particular normative reason in question. The presence of the phrase does not contribute to the proposition expressed by any of (4)-(6), on the factive reading. Rather, (4)-(6) involve a mere stylistic variation on (1)-(3), and indeed on non-psychologistic rationalising statements.

There is, on the other hand, a non-factive reading of (4)-(6) which is liable to be confused with the factive reading. On the non-factive reading of (4)-(6) they tell us that the subject *takes* p to be a normative reason and ϕ s accordingly, and they remain neutral on whether p is a genuine normative reason. On this reading, to say of p that it is *the subject's own reason* is not to say that it is a normative reason which the subject has elected as that for which they will ϕ . Rather, it is to say that it is that which the subject is *taking to be* a normative reason, so that were they asked for their reason for ϕ -ing they would supply their interlocutor with p, and not with some other consideration. And on this reading, (4)-(6) report the same neutral type of rationalising explanation as reported by psychologistic

rationalising statements. If we read (4*) non-factively, what it's saying is that Smith is walking to the quad for what he takes to be a reason in favour of his doing so: that there is free coffee available there. *Smith's own reason for action*, although perhaps not really a reason, is that there is free coffee available in the quad.

There is an interesting feature of the non-factive reading of (4)-(6) which is worth making explicit. On the non-factive reading, there is no concept of a reason applied to anything by the endorsement of (4)-(6). For on the non-factive reading, the sense in which p is S 's reason is that it is something which S takes to be a normative reason. It is not something which is in any sense a *reason*. Read non-factively, to say of p that it is S 's reason for ϕ -ing is not to say that it is a reason for which S ϕ s. On the non-factive reading, then, (4)-(6) really do differ in meaning from (1)-(3) – we've just arrived at independent proof of that. That's because (1)-(3) certainly do apply a concept of a reason to the relevant item at issue.

Not only do I think it obvious that (4)-(6) can be read in the non-factive way described, but I think something of an argument can be put forward for the claim that that's so. Consider what we want to get at when we ask what a subject's values are, when we ask, for example: what are Smith's values when it comes to coffee? A correct answer to that question consists in listing *what Smith values* about coffee: a strong aroma, a mild caffeine kick, a certain amount of bitterness, and so on. The correct answer to the question does not consist in identifying features of coffee which are in fact valuable, where Smith recognises the relevant features to be so. Rather, the correct answer to the question consists in identifying what Smith himself values about coffee: what it is about coffee which Smith *takes to be* valuable. More generally, identifying the agent's values is not a matter of identifying what is valuable and then going on to determine what amongst the correct set of values the agent manages to recognise as so. Rather, it is a matter of identifying what the agent takes to be valuable – what the agent values herself. Talk of S 's values, then, should be read 'non-factively'. This, I suggest, confers plausibility on to the suggestion that talk of S 's reasons can be read non-factively too. If the operator ' S 's x ' functions non-factively in the context of values, we should expect it to be capable of performing the same function with respect to reasons.

The failure to spot that (4)-(6) can be read so that they do not mean the same as (1)-(3) is common in the literature, it is not just Dancy who makes the mistake. To pick but one example, here's the mistake made by Setiya (2011):

...let us concentrate on the explicit articulation of the agent's reasons, as follows:

A is doing ϕ on the ground that p ; that is his reason for doing it.

(*Ibid.*:132)

The phrase which appears before the semi-colon in the sentence Setiya focuses on is an instance of (3a). The phrase which appears after the semi-colon is an anaphoric instance of (4). Setiya shows no sign of recognising that the two sorts of sentences can come apart in meaning.⁵

We've seen that (4)-(6) admit of a factive and of a non-factive reading. It should now be clear where Dancy's argument goes wrong. Either we read (4)-(6) factively or we read them non-factively. If the former, then they do mean the same as (1)-(3), but we cannot intuitively conjoin the denial of *p* to them with coherence. If the latter, then we can do so, but they do not mean the same as (1)-(3), so no conclusion to the effect that (1)-(3) should also be read non-factively follows. I suggest that Dancy has conflated the factive reading with the non-factive reading of (4)-(6), thus generating the illusion that they *both* are to be read non-factively *and* have the same meaning as (1)-(3), when there is, in fact, no single reading of (4)-(6) on which they have both of those properties.

Dancy might respond to all this by saying that I've misread his reasoning. There was never supposed to be an *argument* for the non-factive reading of (1)-(3), he might say, instead the reasoning was supposed to be a naked appeal to the intuition that the denial of *p* can coherently be conjoined by (1)-(3). Thus, (4)-(6) do not really need to be relied upon at all.

In response to this it should be noted that even if one does find it intuitive that (1)-(3) can be coherently conjoined with the denial of *p*, what's been said here provides us with the material to explain away the intuition consistently with the truth of the factive reading of (1)-(3). (4)-(6) admit of two readings: a factive reading and a non-factive reading. On the former, they mean the same as (1)-(3). On the latter, they don't, because they don't apply a concept of a reason to anything. To explain away the intuition that (1)-(3) can coherently be conjoined with the denial of *p* we just have to say that the intuition results from the conflation of (1)-(3) with (4)-(6) where the latter are read non-factively, and that one can be prone to make that mistake because (4)-(6) have an alternative reading on which they do mean the same as (1)-(3).

Alternatively, Dancy might respond to what I've been arguing here by abandoning his reliance of (4)-(6) and instead appealing to another of his favoured constructions:

(7) *S* ϕ ed/is ϕ -ing in the light of *p*

Dancy could then argue that (7) can coherently be conjoined with the denial of *p*, assert that (7) has the same meaning as (1)-(3), and thus conclude that the non-factive reading of (1)-(3) is correct.

In response to this I want to suggest, again, that (7) is ambiguous. (7) can be read factively, so that it has the same meaning as (1)-(3) and indeed as non-psychologistic ra-

⁵I am not the first to notice that (4)-(6) admit of a non-factive reading. Stout (2004) also notes it, although he doesn't develop the point very much and doesn't offer a positive account of what the non-factive reading amounts to.

tionalising statements. But, and I admit that I have to strain my ears here slightly, it can be heard in a non-factive way too. On the non-factive reading, the ‘in the light of’ construction serves as a metaphor for a relation between the subject and what they believe, namely *p*. The meaning of the metaphor is that *p* is said to be something which *appears* to be a normative reason to the subject – *p* is cast in the light of normative reason-hood from the subject’s own point of view, so to speak. Read in that way, (7) means the same as (4)-(6) read non-factively and so cannot mean the same as (1)-(3). Thus, this fresh argument of Dancy’s which appeals to (7) fails in the same way as his Master Argument. We can conclude that the factive reading of (1)-(3), which is the default view, escapes unscathed from Dancy’s attack.

3.3 A Dialectical *Impasse*?

The conclusion just drawn might seem too quick. One might think that even if it has been shown that the material from Dancy can be handled readily by the proponent of the factive reading of (1)-(3), the material still generates a certain problem for the proponent of the dialectical strategy I am pursuing. I now want to give a separate treatment of the dialectical worry at issue here. I’ll describe the issue first, before offering two responses to it.

The dialectical strategy I am pursuing comes in two stages. The first is to demonstrate that my preferred view should be the default view. This is achieved by the eliciting of intuitions in favour of the preferred view. The second stage is to demonstrate that the strongest arguments for the contrary view can be handled by the proponent of my own view. Sometimes the argument for the contrary position will take the form of pointing to a datum and then arguing that only the proponent of the alternative view can handle the datum. Responding to this type of argument would involve showing that, in fact, the proponent of the preferred view can handle the datum just as well or else that the datum isn’t really a datum after all.

Dancy’s Master Argument is an instance of the sort of strategy at issue. The datum in question is linguistic: it is that (4)-(6) do not require the truth of *p*. It is then argued that only the non-factive reading of (1)-(3) can handle that fact, because (1)-(3) mean the same as (4)-(6). My response to this is effectively to suggest that Dancy’s datum isn’t really a datum, for the only datum in this area is weaker than advertised: it is that (4)-(6) admit of a reading on which they do not require the truth of *p*. Thus, Dancy’s Master Argument would require that *on that reading*, (4)-(6) mean the same as (1)-(3). At this point it is suggested that when they are read as not requiring the truth of *p*, they precisely do not mean the same as (1)-(3) for the only way of reading them so that they do not require the truth of *p* requires that those sentences do not apply a concept of a reason to anything, as (1)-(3) do.

Here’s the dialectical worry I want to address. Given that Dancy’s Master Argument

works in the way described, the dialectical strategy I'm pursuing begins to look ineffective. For, it might be protested, all I've done is elicit an intuition in favour of my preferred view and then shown that, on the assumption that my preferred view is correct, Dancy's supposed datum can be handled. But Dancy, it seems, can pursue the very same dialectical strategy in reverse. In particular, Dancy could say that it is indeed a datum that (4)-(6) admit of a single, non-factive reading. He could then say that on this single reading, they mean the same as (1)-(3) and hence conclude that the latter do not require the truth of p. He could then come up with some way of handling my datum: that (1)-(3) mean the same as the non-psychologistic 'because' analogues, from the standpoint of his own non-factive view. Either he could agree, assuming that his non-factive view is correct, that there's grounds for doubting that the datum is really a datum, or else show that his theory can handle it anyway. The resulting state of play would be one in which each party can elicit an intuition in favour of their own position and, from the standpoint of their supposedly intuitive position, handle the opposing parties' datum. The result would be a dialectical *impasse*. The worry is that I've landed myself precisely in that situation.

How to respond to this dialectical worry? One way of responding is to demonstrate that the worry misconceives the way I've attempted to handle Dancy's datum. I haven't merely shown that *supposing the factive reading of (1)-(3) true*, then Dancy's datum can be handled in the way described. On the contrary, the handling of the datum in the way described is compatible with denying that the factive reading of (1)-(3) is true. The handling of the datum is this: (i) there is *a reading* of (4)-(6) on which they do not require the truth of p – that is the only plausible datum in this area; (ii) on that reading the phrase 'S's reason/basis/grounds' means 'what S takes to be a normative reason'; (iii) given (ii), it follows that (4)-(6) do not apply a concept of a reason to anything on the only non-factive reading thereof, and hence do not mean the same as (1)-(3), which do. All of that is supposed to be plausible independently of whether one goes for the factive reading of (1)-(3). Thus, Dancy's datum, and the Master Argument associated with it, is no good, even if one agrees with him about (1)-(3).

So there isn't a dialectical *impasse* generated in the way described because I am not in the business of eliciting an intuition in favour of my preferred view and then showing that *from the standpoint of my own position*, the datum, if indeed it is one, is handle-able. Rather, I'm in the business of eliciting the initial intuition and then showing that *everyone* should reject Dancy's datum in favour of treating a weaker claim as a datum, where it's also independently plausible that that weaker claim doesn't really get Dancy what he wants. Even if Dancy does come up with a way of handling my datum from the standpoint of his own position, then, that would not generate an *impasse* in the way described, because his argument has been shown to be independently implausible.

But there is a second way of handling the dialectical worry which is stronger than the one just broached, for it grants that the description of my strategy is correct, but goes on to

simply deny that Dancy does have a good way of handling my datum. We can begin with the question: how does Dancy propose to handle ‘S ϕ ed/is ϕ -ing because p’ talk? Here’s a statement of his position:⁶

There are factive explanations of action in terms of the agent’s reasons. For instance, we have the non-factive:

His reason for doing it was that it would increase his pension.

But we have factive ways of saying the very same thing, for instance:

He did it because it would increase his pension.

I take it that unlike the former, the latter cannot coherently be continued ‘but he was sadly mistaken about that’, which is the mark of its factivity. Should we say, then, that we have here two explanations of the same thing, two ‘rational’ or rationalizing explanations, one of which is factive and one of which is not? In general, such a situation is no cause for complaint. But I think that the explanation itself is not a factive one; there are just factive ways of wording it. The word ‘because’ renders the explanation factive, but only in the sense that it commits the explainer to the truth of the explanans; the explanation itself, which might have been given in other terms, remains stubbornly non-factive. That is, the fact that the explanation can be given in non-factive form shows that it is a non-factive explanation, even though that very explanation can be given in ways that (for trivial reasons to do with the use of certain words) are themselves factive. In the latter cases, it is not the explanation itself that is factive, but the form of words that we use to express it. (Dancy, 2004b:28)

In the above passage Dancy focuses on the relationship between an instance of (4) and a corresponding instance of a non-psychologistic rationalising statement. Since he takes (4)-(6) to always have the same meaning as (1)-(3), the passage tells us something about his views on the relationship between (1)-(3) and non-psychologistic rationalising statements. Thus, in interpreting this passage, I will pretend that he is comparing an instance of (1)-(3) to a non-psychologistic rationalising statement.

The view seems to be this. (1)-(3), just like (4)-(6) and (7), report the presence of a certain kind of rationalising explanation. The rationalising explanation does not require the truth of p, but takes p as *explanans*. It is hence a type of explanation why that is non-factive and non-causal, which is how Dancy thinks of rationalising explanations. That

⁶Although the passage does not come from *Practical Reality* I’m confident that the view expressed in it matches the view expressed in that book.

those sentences report the obtaining of the relevant sort of explanatory state of affairs exhausts their meaning.

Now, Dancy agrees with what I want to treat as a datum: that non-psychologistic rationalising statements say the same thing as (1)-(3). He does not wish to deny that my datum is really a datum. So what he wishes to say is that non-psychologistic rationalising statements report the same non-factive, non-causal, non-psychologistic rationalising explanation as their (1)-(3) counterparts and that that exhausts their meaning. Such statements, then, although they mean the same as (1)-(3) don't really report the success condition of responding to a normative reason. Rather, they just report the obtaining of a rationalising explanation, as Dancy thinks of it.

Dancy's view of non-psychologistic rationalising statements, however, generates an obvious problem: 'S ϕ s because p' requires the truth of p – such statements are factive. How does Dancy propose to render this consistent with his own account of such statements? What he says is effectively that 'because' talk performs a merely pragmatic function: it functions to make it explicit that the speaker wishes to commit themselves to the truth of p, on top of asserting the explanatory proposition expressed by the statement in question. The thought expressed by the non-psychologistic rationalising statements is the same as the thought expressed by (1)-(3), it's just that the former involve an idiom which performs the merely pragmatic function of evincing the speaker's commitment to p.

So Dancy would say, in response to me, that it's agreed that (1)-(3) mean the same as their 'because' counterparts. It's just that the latter aren't really statements that report the success notion. They are factive, but their factivity is an artefact of the pragmatic function performed by the 'because' which appears in them. Thus, my datum can be handled by Dancy's theory.

However, Dancy's account of 'S ϕ ed/is ϕ -ing because p' talk is, it seems to me, inadequate for a number of reasons. First, as he himself seems to acknowledge, there is such a thing as ϕ -ing in response to a normative reason: there is such a thing as ϕ -ing for a good reason. If we grant that, then presumably we must also grant that English affords us a way of talking about such a phenomenon. But what way could that be other than: 'S ϕ ed/is ϕ -ing because p', in the rationalising sense of 'because'? On Dancy's view of such sentences, it is not true that they are used to report the success condition: the thought we have in mind when we assert 'S is ϕ -ing because p' is not that S is ϕ -ing in response to the normative reason that p, but that one of Dancy's non-causal, non-factive, non-psychologistic rationalising explanations holds true of the subject's ϕ -ing. We are also, incidentally, evincing a commitment to p and we are doing so by exploiting a device in English which, in the context in question, functions solely to enable us to do so. We are not saying, of p, that it is a normative reason to which S is responding by ϕ -ing. But to repeat: if the 'S is ϕ -ing because p' idiom doesn't afford us a way of saying that, then what does?

Another problem with Dancy's account can be brought out by considering, once again,

(1a*):

(1a*) Smith is walking to the quad for the reason that there is free coffee available there

If I were to assert (1a*) but add to it: ‘but there is no free coffee available in the quad’ I would end up asserting something on the face of it incoherent.⁷ Of course, on my preferred view, we should take the incoherence at face value, for appending the denial of p to ‘S ϕ s for the reason that p’ is in fact contradictory. But my opponent will have to acknowledge and account for the oddness as well. It’s just that they’ll have to deny that (1a*) strictly speaking entails that there is free coffee available in the quad. Instead, they’ll have to say that that fact is presupposed by the speaker who asserts (1a*), but where the presupposition in question is not also entailed by the sentence asserted.⁸

But now suppose Smith were himself to assert a first-person variant of (1a*):

(1a_{fp}) I am walking to the quad for the reason that there is free coffee available there

There would be a kind of incoherence which would attach to Smith’s assertion of (1a_{fp}), were he to append to it: ‘but there is no free coffee available in the quad’ which is not present in the corresponding third-person case in which that denial is added. This, I submit, is because (1a_{fp}) not only involves the presupposition (whether entailed or not) that there is free coffee available in the quad, but also involves *Smith expressing his belief that there is free coffee available in the quad*. It involves this by dint of its first-person form. Thus, combining (1a_{fp}) with the denial at issue would involve a kind of incoherence at least analogous to the incoherence exemplified by instances of the belief-ascription sentences that form the basis of Moore’s Paradox: ‘I believe that p but not-p’.

But now let’s suppose that Smith were to move from asserting (1a_{fp}) to asserting:

(1_{bc}*) I am walking to the quad because there is free coffee available there

In my view, (1a_{fp}) and (1_{bc}*) mean exactly the same thing, and on Dancy’s view likewise. But on my view each of these are mere stylistic variants on the other. For Dancy, the difference is more substantial: there is a pragmatic feature of (1_{bc}*) which (1a_{fp}) does not exemplify. The former is supposed to carry with it a device which serves to evince Smith’s belief that there is coffee available in the quad. However, that device would serve no function that isn’t already guaranteed to be served by the first-person form of the latter. Once we focus on the relevant first-person forms, then, Dancy’s account becomes dubious for he is committed to saying that there is a difference between the ‘because’ idiom and

⁷It should be noted that the same thing does not hold of the relevant instance of (4), on the non-factive reading. It is of course this fact which Dancy’s Master Argument exploits in order to generate the implausible result that (1)-(3) don’t record the success condition.

⁸See Comesana and McGrath (2014) for a development of such an account.

the (1)-(3) idioms, when those are utilised first-personally, which cannot be there. That's so even though the difference is supposed to be merely at the pragmatic level.

So Dancy's account of 'S ϕ ed/is ϕ -ing because p' talk is objectionable. Thus, Dancy does not have an adequate way of dealing with my datum – that such talk means the same as (1)-(3) – from the standpoint of his own theory. The dialectical *impasse* worry, according to which Dancy can simply adopt my dialectical strategy in reverse with propriety, can hence be avoided in more than one way. First, it can be avoided by appeal to the point that my dealing with Dancy's datum is independent of what verdict we pass on (1)-(3). Second, it can be avoided by appeal to the point that Dancy does not have an adequate way of dealing with my datum himself.

3.4 The Argument from Rationality Rejected

In this section I want to examine, and ultimately reject, a second argument for the non-factive reading of (1)-(3) which has been advanced, separately, by Schroeder (2008) and by Comesaña and McGrath (2014). The argument appeals to the thought that the agent ϕ s rationally in the bad case, and then infers from that that the agent ϕ s for a reason in the bad case. Schroeder's argument differs in minor detail from Comesaña and McGrath's. For example, Schroeder doesn't invoke the distinction, which I'll get to in a moment, between substantive rationality and structural rationality. Instead, Schroeder operates with an undefined notion of rationality. I'll be focusing on Comesaña and McGrath's version of the argument for, given that they draw that distinction, their argument is better developed. But my response to their argument is just as effective against Schroeder's variant of it.

Comesaña and McGrath's primary interest is not the phenomenon captured by (1)-(3) of ϕ -ing for a reason, but the related phenomenon of *possessing* or *having* a reason to ϕ . Roughly, that notion amounts to the notion of a psychological link between the subject and p, so that p is readily available to them as the reason for which they ϕ . The argument I'm going to investigate draws the sub-conclusion that S can ϕ for the reason that p in the bad case, and then moves from there to the thought that S can possess the reason that p in the bad case, which is Comesaña and McGrath's ultimate conclusion. Since I'm more interested in ϕ -ing for a reason I'll restrict my attention to what consequences the argument has for that phenomenon and ignore the further conclusions drawn concerning the related phenomenon of possessing a reason.⁹

First of all we need the distinction between the two notions of rationality already alluded to. The distinction is, to borrow some terminology from Scanlon (2003), between *substantive* rationality and *structural* rationality. We can bring out the contrast here by focusing on the question: what do we mean when we say, of an agent, that they *ought* to

⁹Schroeder (2008, 2011) is also interested primarily in the phenomenon of possessing a reason but, again, his argument has application with respect to the notion of ϕ -ing for a reason too.

ϕ ?¹⁰ One thing we might mean, as we've already seen, is that there is decisive reason for the agent to ϕ . This is a matter of a certain sort of relation holding between the subject's ϕ -ing and the world: a relation of decisively favouring which links a certain set of facts to one's ϕ -ing. On the other hand, when we say that someone ought to ϕ we might have something seemingly quite different in mind: that *given that they have certain (other) attitudes*, it's rational for them to ϕ , which is to say that, given that they have the relevant attitudes, it would be irrational for them not to ϕ . This second sort of ought is a matter of a certain relation holding between an attitude, or set of attitudes, one has and one's ϕ -ing (where ϕ -ing can be another attitude available to one, or else an act or omission available to one). We determine when one ought to ϕ in this second sense in abstraction from what facts in the world constitute what can be said for and against the agent's ϕ -ing. Instead, we focus solely on the attitudes the agent has.

These two notions are genuinely distinct. There might be conclusive reason for one to ϕ even if one has no set of attitudes such that it would be irrational for one not to ϕ . For example, unbeknownst to me my car might have a flat tyre. That the car has a flat tyre is a decisive reason for me to change it. But I might have no idea that the car has a flat tyre. Thus, I have no attitude such that it would be irrational for me not to change it. Likewise, I might have an attitude such that it would be irrational for me not to ϕ , even though there is nothing to be said in favour of my ϕ -ing. I might believe, mistakenly, that I live on a fault-line. Given that I believe that, and that I believe that houses built on fault-lines are less stable than houses which are not, I ought to believe that my house is less stable than those not built on a fault-line. But in fact I am wrong and there isn't really any reason at all for me to believe that my house is built on a fault-line. So it would be rational for me to believe that my house is less stable than others, even though it's not the case that there is anything to be said in favour of my thinking that.

There are a great many general principles which tell us when one ought to ϕ in the second sense.¹¹ Examining a handful of them might help to give us a clearer picture of what the current distinction amounts to. Take, for example, the following set of principles:

Means-Ends If S desires to ψ and believes that ϕ -ing is necessary for ψ -ing, then S ought to ϕ

Modus Ponens If S believes that p, and S believes that if p then q, then S ought to believe that q

C+ If S believes that there is conclusive reason to ϕ , then S ought to ϕ

C- If S believes that there is conclusive reason for them not to ϕ , then S ought not to ϕ

¹⁰My presentation of the distinction here closely follows that of Kolodny (2005).

¹¹It might be that there are only a small sub-set of these principles which constitute the fundamental ones – those by dint of which the truth of the rest can be explained. Kolodny (*ibid.*) thinks that the C+ and C- principles I'm about to state occupy such a position.

In each case, what we have is a principle which tells us that in circumstances in which one possesses the attitudes specified in the antecedent, a certain ought applies to one – our second sort of ought. The ought in question applies to one independently of what can be said for and against one's ϕ -ing, prior to one having the relevant attitudes. The ought holds just by dint of one's possessing the attitudes in question.

The first sort of ought is the ought of substantive rationality, the second the ought of structural rationality. There are, of course, many lively debates about quite how to understand the relation between substantive and structural rationality. For example, there's the issue of whether we should think that rational requirements generate normative reasons for one to ϕ in response to them. And if so, what we should identify such reasons with and how we are to enable this suggestion to avoid problematic bootstrapping.¹² We need not get involved with those debates here.¹³

There is a further background distinction which needs to be drawn, one which has a familiar Kantian flavour. This is the distinction between ϕ -ing *in accordance* with what rationality requires of one, and ϕ -ing *rationally*. This distinction cuts across the distinction just drawn between substantive and structural rationality to yield us two separate distinctions. On the one hand there's the distinction between ϕ -ing in accordance with what substantive rationality requires of one and ϕ -ing in a substantively rational way. This is a matter of ϕ -ing in a way that happens to be decisively favoured by the reasons there are for one to ϕ and ϕ -ing in response to those reasons. That is just the Kantian distinction I started out with, but where the normative reasons in question are decisive. On the other hand, there's the distinction between ϕ -ing in a way that's required by some principle of rationality which, given the antecedent attitudes one has, applies to one, and ϕ -ing in a way that is not merely in accordance with what structural rationality requires but is a matter of ϕ -ing for the sake of what structural rationality requires. This is just an analogue of the Kantian distinction I started out with applied to structural rationality.

We have a rough-and-ready intuitive notion of ϕ -ing in response to a normative reason,

¹²See Kolodny (2005, 2007), Broome (1999, 2007), Dancy (2000) and Raz (2011) for a sample of this debate.

¹³As I have said, when one ought to ϕ in the first sense, we can say that it is *substantively rational* for one to ϕ . When one ought to ϕ in the second sense we can say that it is *structurally rational* for one to ϕ . The latter label is appropriate because rationality in the second sense is a matter of the avoidance of *incoherence* between one's attitudes and one's ϕ -ing. The distinction I have in mind is not always referred to using the Scanlonian labels I favour. Kolodny (2005) labels the distinction the distinction between the *ought of normative reasons* and the *ought of rationality*, and he refers to the factors which generate the first *reasons* (in the sense of normative reasons) and the sorts of principles associated with the second *requirements*, or *requirements of rationality* where that terminology is in turn borrowed from Broome (1999). Another appropriate label might be the distinction between *objective* and *subjective* rationality, which Kolodny himself also sometimes uses. Hooker and Struemer (2004) call structural rationality *procedural* rationality. I use the Scanlonian labels because those are the labels favoured by Comesaña and McGrath. Whatever labels we decide to use is immaterial. The point to bear in mind is that there is an important difference between it being the case that one ought to ϕ by dint of there being decisive reasons in existence for one to ϕ and it being the case that one ought to ϕ because one is in some state of mind such that it would be incoherent for one not to ϕ .

and hence ϕ -ing in a substantively rational way. That intuitive notion has been brought in to play numerous times so far. But the notion of ϕ -ing in a structurally rational way is, I think, quite opaque. What does ϕ -ing in a structurally rational way amount to? My own view on the matter is this. To ϕ in a structurally rational way is not to ϕ for the reason *that it is structurally rational for me to ϕ* , where that is taken to be a normative reason for one to ϕ . Nor is it to ϕ for the reason *that I am in M*, where M is the attitude or set of attitudes that correspond to the antecedent of the relevant rational requirement, and where that fact is taken to be a normative reason for one to ϕ . Neither of these suggestions is plausible: we would judge an agent as ϕ -ing in a structurally rational way even if we know that they don't have any conception of the rational requirement which applies to them.¹⁴ Instead, ϕ -ing in a structurally rational way is just taking something to be a normative reason and ϕ -ing accordingly, so that a neutral rationalising explanation is true of one. The relevant consideration is the content of the belief which corresponds to the antecedent of the rational requirement that one is ϕ -ing for the sake of: that ϕ -ing is necessary for ψ -ing, that p and that if p then q, that there is conclusive reason to ϕ , that there is conclusive reason not to ϕ , and so on. ϕ -ing in a structurally rational way is, in a certain sense, transparent to ϕ -ing for a reason: to ϕ in a structurally rational way just is to ϕ in response to apparent normative reasons (if not a genuine one).

I'm now in a position to state Comesaña and McGrath's argument. They rely on the following principle connecting ϕ -ing in a rational way with ϕ -ing for a reason:

(RR) One does something rationally only if one has reasons that make it reasonable for one to do it and one does it on the basis of some (sub-) set of those reasons, i.e. one does it "for" those reasons. (Comesaña and McGrath, 2014:62)

(RR) is intended to be a claim about *substantive* rationality. Given that, (RR) is plausible. Substantive rationality is a matter of there being decisive reasons in existence for one to ϕ . To ϕ in a substantively rational way, then, must be to respond to those factors which make it substantively rational for one to ϕ : to ϕ in response to the relevant normative reasons. When one responds to the normative reason that p, (1)-(3) is true of one – no one doubts that claim. So when one ϕ s in a substantively rational way, (1)-(3) are true of one: one ϕ s for a reason.¹⁵

They then focus on the following pair of cases, variants of the famous case of Bernie

¹⁴See Kolodny (2005:§4) for an argument for the claim that, where beliefs are concerned at least, it is not possible for ϕ -ing in a structurally rational way to take such forms.

¹⁵Comesaña and McGrath derive (RR) from two plausible looking principles (Comesaña and McGrath, 2014:61). The first says that it is substantively rational for one to ϕ only if there is a reason for one to ϕ , where that reason makes it rational for one to ϕ . The second says that one ϕ s substantively rationally only if there is something which makes it rational for one to ϕ and one ϕ s in the light of that factor. I leave specification of these extra principles, and indeed others that I discuss in the following three footnotes, out of the main body of the text in order to avoid an unhelpful and unnecessary proliferation of principles and theses.

and the gin and tonic introduced by Williams (1980). First, there's the following good case:

Thirsty Bernie: Bernie is at a party and is in the mood for a drink. At the bartender's counter, there are glasses of gin and tonic and bottles of a good Belgian ale. Bernie reaches for a glass of gin and tonic, remembering the host's superb gin and tonics, and preferring them even over a good Belgian ale. The glass does indeed contain gin and tonic. All goes well. (Comesaña and McGrath, 2014:62)

Then there is the following bad case:

Deceived Bernie: Bernie is at a party and is in the mood for a drink. Everything is the same as in Thirsty Bernie except that on the bartender's counter the glasses are not filled with gin and tonic, but only tonic (the bartender forgot to add gin). Bernie reaches for a glass just as in Thirsty Bernie. In a moment, Bernie is disappointed. (*Ibid.*)

Comesaña and McGrath suggest take it as a datum that Deceived Bernie ϕ s in a rational way.¹⁶ Given (RR) it follows that Deceived Bernie ϕ s for a reason: (1)-(3) is true of him. This leads us the question: what is the reason for which Deceived Bernie ϕ s? Well, the reason for which Thirsty Bernie reaches for the glass is that the glass contains gin and tonic. If the factive reading of (1)-(3) is true, then that cannot be Deceived Bernie's reason, for it is false in his situation. What else could his reason be? Comesaña and McGrath go through the three most plausible options and dismiss each of them: (i) that he believes that there is gin and tonic in the glass; (ii) that it is probable that the glass contains gin and tonic; and (iii) the set of propositions which make it probable that the glass contains gin and tonic (for example: that the glass looks to contain gin and tonic, that the bartender is not generally forgetful, and so on). In each case, the candidate reason is dismissed on the grounds that Deceived Bernie is the same in all relevant psychological respects to Thirsty Bernie.¹⁷ The thought here, I take it, is that what Thirsty Bernie takes to be a reason for his ϕ -ing is the same as what Deceived Bernie does, so that whatever consideration the latter ϕ s on the basis of, if any, must be the same as that which Thirsty Bernie ϕ s on the basis of.¹⁸ Since the only remaining option is to agree that the reason for which Deceived

¹⁶Comesaña and McGrath (2014:62). They call that claim the *Rationality Assumption*.

¹⁷They dismiss (i)-(iii) for other reasons too. For example, they think (iii) violates a further plausible thesis: that Deceived Bernie and Thirsty Bernie are *equally* rational in ϕ -ing.

¹⁸Comesaña and McGrath (*ibid.*:62-63). They call the claim that the two subjects are the same in all relevant psychological respects the *Sameness of Psychological Basis Assumption*. Their formulation of the assumption is, however, confusing: "The second assumption is that the psychological story about Bernie's basis for reaching for the glass is the same in the two cases: the same consideration moves Bernie to act in the two cases." What's said here is confusing for two reasons. First, what follows the colon is not obviously equivalent to what comes before the colon. It might be that there are psychological differences between the two cases (for example: Thirsty Bernie *knows* but Deceived Bernie doesn't) even though the

Bernie reaches for the glass is that the glass contains gin and tonic, it follows that one can ϕ for the reason that p even though p is not true. The factive reading of (1)-(3), then, is false.

The problem with this argument is the claim that Deceived Bernie ϕ s in a rational way. Since (RR) concerns substantive rationality, the sense in which Deceived Bernie ϕ s rationally needs to be that he ϕ s in a substantively rational way, if the principle (RR) in particular is to generate for us the result that (1)-(3) is true of Deceived Bernie. But S ϕ s in a substantively rational way only if there is a *normative reason* for S to ϕ , and S ϕ s in response to it: (RR) is a true principle which connects a certain notion of rationality with the notion of a *normative reason*. But Bernie does not ϕ in response to a normative reason. Rather, there is an apparent reason on the scene and he ϕ s accordingly. The apparent reason is that the glass contains gin and tonic. But as a falsehood, that proposition can be nothing more than a *merely* apparent reason. So Deceived Bernie does not ϕ in a substantively rational way. Given that, (RR) is useless, when applied to the case of Deceived Bernie, in generating the result that (1)-(3) is true of Deceived Bernie.

It is indeed a datum that Deceived Bernie is ϕ -ing rationally – Comesaña and McGrath have got that much correct. But the only sense in which Deceived Bernie is ϕ -ing rationally is that he is ϕ -ing in a *structurally rational* way. That is to say, he is ϕ -ing in a way that is both in-keeping with a rational requirement that applies to him, and he is ϕ -ing in the light of the relevant rational requirement. The rational requirement in question is this: if one believes that p, and one takes p to be a conclusive reason for one to ϕ , then one ought to ϕ .¹⁹ As I have said, I think that ϕ -ing in a structurally rational way reduces to ϕ -ing in a way that manifests one's taking there to be a decisive reason in favour of one's ϕ -ing. So the correct thing to say about Deceived Bernie's rationality is that he reaches for the glass in a way that manifests his taking there to be a conclusive reason for him to reach for the glass, so that he satisfies the rational requirement just specified.

In sum, then, we can say of Deceived Bernie that he ϕ s rationally. We can say that he ϕ s for an apparent normative reason. We can say that Bernie reaches for the glass because he believes that the glass contains gin and tonic, in the rationalising sense of 'because'. We can also say that (4)-(6) and (7) are true of him on the non-factive readings thereof: *Bernie's reason*, in the case in question, is that the glass contains gin and tonic, and there is a sense in which Bernie ϕ s in he light of the glass containing gin and tonic. But Bernie

consideration on the basis of which they act is the same. Second, it's confusing because describing the case of Deceived Bernie as a case in which there is a consideration or basis for which Deceived Bernie acts is potentially question-begging, for the proponent of the factive reading will precisely want to say that there is no basis on which Deceived Bernie is ϕ -ing. In the text, I've attempted to avoid these worries by stating the essential point of the second assumption in a different way.

¹⁹Strictly speaking, this principle would have to be modified if it were to stand a chance of being true. It would have to say that if one *rationally* believes that there is conclusive reason for one to ϕ and one *rationally* takes p to be a conclusive reason for one to ϕ , then one ought to ϕ . I don't think this makes a difference to the argument in the text.

does not reach for the glass *for a reason*. (1)-(3) are false of him.

Comesaña and McGrath might respond to this by saying that it is the case that Bernie ϕ s in a substantively rational way and that does indeed require that Bernie ϕ s in response to a normative reason. But *normative reasons needn't be facts*: whether p is true or false doesn't affect its status as a normative reason; (NR=F) is false. One might read the following passage as intended to assert such an idea:

There are many, many considerations out there for and against any action. We care about finding out which ones are true, since those are the ones that will make our lives go best. Still, the other ones are still reasons, and if someone acts on them, their life might not go best. . . , but they will be acting rationally. (Comesaña and McGrath, 2014:78)

One could read this passage as saying that p, whether true or false, is a reason in favour of one's ϕ -ing. What truth adds to p is not a different status reasons-wise, but the following property: ϕ -ing for the reason that p is conducive to the successful achievement of one's goals. If p were false, then it wouldn't have that property. That's why it's important to us that we ϕ only for sake of true considerations. But, to repeat, its having that status isn't required for it being a normative reason.

If this were right, then our conception of the dialectic being addressed in the current chapter would have to change significantly. For if it were true that normative reason needn't be facts, then we'd have to separate the idea that the notion of reason involved in (1)-(3) is a normative reason, so that (1)-(3) can be thought of as recording the success phenomenon of responding to a normative reason, from the idea that (1)-(3) are factive in the sense of requiring the truth of p. The claim that (1)-(3) requires the truth of p would no longer be entailed by the claim that (1)-(3) express the success notion.

It is, however, implausible that normative reasons can be falsehoods. There are two reasons for endorsing (NR=F) and a challenge to those who deny it, which I now want to draw attention to.

First, if p is a normative reason, then p is something which can explain why one pro-tanto ought to ϕ : p is a normative reason only if 'S (pro-tanto) ought to ϕ because p' is true.²⁰ But p can explain why q only if p is true. Explanation why is factive. Hence: p is normative reason only if p is true.

Second, I noted in Chapter One that one can ϕ in response to the normative reason that p (and more generally: can ϕ in a way that's subject to a rationalising explanation), only if p is present to mind in the right sort of way, which involves p being present to mind *qua* normative reason. (DT) is part of the truth here, in my view: for p to be present to mind in

²⁰This is not to endorse the claim, defended by Broome (2004), that what it is to be a normative reason is for it to play that explanatory role. The claim which I defend in the text is weaker than that: it is merely a necessity thesis, not a thesis to the effect that the explanatory status is constitutive of the normative status.

the right sort of way requires that p is *believed*. But even if one denied (DT) in favour of a more liberal thesis so that, for example, sensory experience could constitute the presence to mind of p in the right sort of way, still it seems uncontroversial that only *truth-directed* attitudes can constitute the presence to mind of reasons in the way at issue. Desiring that p , hoping that p , wishing that p , suspending judgement about whether p , and so on, cannot constitute the presence to mind of p in the right sort of way. Only attitudes which one has towards p only if p is thereby *taken to be true* are candidates to do the relevant work. Only cognitive attitudes – sensory experiences (assuming that they have content), beliefs, articles of knowledge and the like – can play the role, and even then I think it's only the *commitment* attitudes like belief and knowledge which can do it. But now let's ask the question: why does ϕ -ing in response to the normative reason that p require that one has some truth-directed attitude towards p ? A plausible answer here, I suggest, is that p is a normative reason only if p is true. If it's part of what it is to be a normative reason that the relevant item is true, then no wonder awareness of normative reasons *as such* needs to be truth-directed awareness.

Finally, the challenge. When one finds out that p is false, one ceases to treat p as a reason in favour of one's ϕ -ing: one is no longer disposed to rely on p in reasoning and one is no longer disposed to offer p as a reason to anyone who asks, for example. But having those dispositions is necessary for taking p to be a normative reason at all. Hence we are disposed to no longer think of p as a normative reason, upon learning that p is false. If Comesaña and McGrath were right, then this attitude would be mistaken. But they have not supplied any independent reason for thinking that it is. The challenge to them is to provide such reason for thinking that we're incorrect to cease treating p as a reason when we find out that it's false.

We can conclude, then, that the Argument from Rationality is unsound. The factive reading of (1)-(3) still stands.

3.5 The Argument from Incredulity Rejected

As I noted above, the factive reading of (1)-(3) implies that the subject in the bad case is not a subject who ϕ s for a reason. Were I to falsely but blamelessly believe that my airport boarding gate is about to close and thus start running towards it accordingly, the factive reading of (1)-(3) would imply that I'm not running for the reason that my gate is about to close. Since that's the only candidate for the reason for which I act, it follows that I'm running for no reason at all. But that is surely a suggestion which should be met with incredulity, is it not?

In response to this worry I think it should be pointed out that there is plenty which the proponent of the factive reading can say about the bad case which I think should assuage the incredulity. The proponent of the factive reading can say that the subject in the bad

case is ϕ -ing because they believe that p, in the rationalising sense of ‘because’; that they are taking p to be a normative reason and are ϕ -ing accordingly; that they are ϕ -ing in a (structurally) rational way; that *their own reason* for ϕ -ing is that p – (4)-(6), on the non-factive reading, are true of them; and that they are ϕ -ing in the light of p, on the non-factive reading thereof too.

Given all that, one should not find it too difficult to allow that the subject in the bad case does not also ϕ *for a reason*. The thought that that result is incredible is driven by the conflation of the claim that the agent is ϕ -ing for no reason with the claim that it is a case in which the agent is not even so much as taking p to be a reason and ϕ -ing accordingly – it is a case of an agent being merely caused to ϕ . But the proponent of the factive reading of (1)-(3) needn’t say that, for they can say that all of the conditions mentioned in the preceding paragraph hold in the bad case.

3.6 The Argument from Transparency Rejected

I turn now to the final argument I want to dispatch against the factive reading. I’ve been exploring the state of ϕ -ing for a reason, and in particular I’ve been exploring an issue in what we can think of as the metaphysics of ϕ -ing for a reason: whether it is possible to ϕ for a reason in the bad case. But there is a certain *epistemic* issue involving that phenomenon which I now want to draw attention to. The epistemic issue is this: when one ϕ s for the reason that p, is one always thereby in a position to know that one ϕ s for the reason that p? And relatedly: when one ϕ s, but one doesn’t ϕ for the reason that p, is one in a position to know that one doesn’t? Let’s codify the two epistemic theses:

Reasons-Luminosity+ If S ϕ s for the reason that p, then S is in a position to know that they are ϕ -ing for the reason that p.

Reasons-Luminosity- If S ϕ s and it is not the case that S ϕ s for the reason that p, then S is in a position to know that they are not ϕ -ing for the reason that p.

Let’s call the conjunction of Reasons-Luminosity+ and Reasons-Luminosity- *Reasons-Transparency*. In a slogan, Reasons-Transparency says that if one ϕ s – if one acts, omits, believes, desires, intends, hopes, wishes, etc. – then one is always in a position to know whether or not one is doing so for the reason that p. For any instance of ϕ -ing one is undertaking, it is transparent to one what is going on with one reasons-wise, with respect to one’s ϕ -ing.

As it stands, Reasons-Transparency is in need of some clarification, if it’s to seem at all plausible. To see why consider the following passage from Dancy (2008):

Its seeming to me as if I am acting in the light of a particular fact is no proof at all that this fact is really the reason for which I am acting. We just don’t have

this sort of access to our reasons. It might well be that I seem to myself to be acting from pure altruism when in fact my motivation is not at all something to congratulate myself about. (Dancy, 2008:271)

As Dancy notes, it is possible for a subject to ϕ for the reason that p , while making a mistake about the reason for which they are ϕ -ing. One might ϕ for the reason that p , yet mistakenly believe that one is ϕ -ing for the reason that q , or for no reason at all. This might happen for a reason internal to the subject. For example, it might happen because of self-deception on the subject's part: perhaps they have some reason for not admitting to themselves that they are ϕ -ing for the reason that p because it would reveal something about their character which they don't want to admit. But it might be a simple case of misremembering or confusion on the subject's part. On the other hand the sort of error might come about for a reason external to the subject. For example, it might happen because were the subject to start reflecting on what their reason is for ϕ -ing, that would trigger into action a hypnotist who has set himself the anti-Socratic task of making sure that people are mistaken about what is going on in their own minds.

Reasons-Transparency is consistent with the possibility of such cases, for the relevant notion of being in a position to know is weak enough to allow for the possibility of them. Someone who makes the relevant sorts of mistakes might still be in a position to know in the sense at issue, they would just not be in a position to exploit their favourable epistemic condition in order to arrive at the truth, because of the relevant internal or external disabling conditions.²¹ Of course, there might be another sense of 'being in a position to know' where the subjects in the cases at issue are not in a position to know. But if so, all that follows is that we shouldn't understand the notion of being in a position to know utilised by Reasons-Transparency in *that* way.

What reason is there to endorse Reasons-Transparency? One might wish to endorse it because one endorses Transparency about every psychological condition, and ϕ -ing for a reason is a psychological condition. A more interesting motivation, however, can be extracted from the following remarks made by Williamson in the context of a discussion about the nature of evidence and our possession of it:²²

Rational thinkers respect their evidence. Properly understood, that is a platitude. But how can one respect one's evidence unless one knows what it is? So must not rational thinkers know what their evidence is? If so, then for rational subjects the condition that one has such-and-such evidence should be non-trivial yet luminous. (Williamson, 2000:164)

There is an argument gestured at in this passage which purports to prove that the evidence possessed by the subject is a luminous condition: that the possession or having of a

²¹See McHugh (2010) for an elaboration and defence of such a weak conception of being in a position to know.

²²It should be noted that Williamson does not himself endorse the argument or arguments in question.

piece of evidence, in the sense of ‘possessing’ or ‘having’ which interests Comesaña and McGrath, satisfies one or both of a Luminosity+ and a Luminosity- thesis. What I want to do now is describe an argument for Reasons-Transparency which can be extracted from the passage from Williamson, with the concept of evidence replaced with that of a reason.

The argument in question is this. ϕ -ing in a rational way is something that it is always in our power to do. If we ever fail to ϕ rationally, that will be because of a mistake on our part which, as long as our cognitive functions are working properly, we aren’t prone to self-deception, and we aren’t in a position where an external source would stop us from being able to exploit an opportunity for knowledge, we can correct for just by reflecting and reasoning more carefully. But if that’s so, then it must be that we are always in a position to tell whether or not we’re ϕ -ing for the reason that p . For suppose we weren’t. Then in the situations in question we might still not be able to know what is the reason for which we believe, if any. But then, how are we to go about changing the rational credentials of our ϕ -ing, with a view to improving it? And if we can’t, then in those circumstances it’s not in our control to ϕ rationally afterall, contradicting our initial assumption. So Reasons-Transparency must be true.

I’ve introduced and explained Reasons-Transparency, and I’ve provided a sketch of two motivations for it. I’m now in a position to state the final argument against the factive reading of (1)-(3) I want to focus on in this chapter. The argument takes as its starting point the thought that Reasons-Transparency is true: when one ϕ s, one is always in a position to know whether or not one is ϕ -ing for the reason that p . But, the argument continues, if (1)-(3) are true only in the good case then Transparency is false. That’s because if the factive reading of (1)-(3) is true, then Reasons-Luminosity- is false. So, (1)-(3) must be true across good and bad cases. Only then can we secure the sort of infallibility about our reasons for ϕ -ing engendered by the supposedly plausible Reasons-Transparency.

Why does the factive reading of (1)-(3) rule out Reasons-Luminosity-? To see why, consider the case of the brain-in-a-vat. Suppose the envatted subject believes that it’s going to rain later on today, because, unbeknownst to them, they’ve been induced to suffer a hallucination of a weather report which told them so. So they believe that it’s going to rain later because they believe falsely that a reliable weather report told them so. If the factive reading of (1)-(3) is true, then they don’t believe for a reason, even though they are not in a position to tell that they don’t believe for a reason. The envatted subject would believe falsely, were they to reflect on their situation, that they believe that it’s going to rain later for the reason that a reliable weather report told them so, and they are not in a position to believe in any other way. So if the factive reading is true, Reasons-Luminosity- is false.

Although the argument relies only on the thought that the factive reading of (1)-(3) rules out Reasons-Transparency by ruling out Reasons-Luminosity-, I think it’s worth noting that the factive reading also serves to undermine Reasons-Luminosity+, too, even

though it is not straightforwardly inconsistent with it. That's because the factive reading is inconsistent with the more interesting way of motivating the Reasons-Transparency thesis identified above. If the factive reading of (1)-(3) is true, then the brain-in-a-vat is not in a position to control for whether they believe for a reason anyway. So that undermines the general thesis that we can always control for the rational credentials of our ϕ -ings, which is one of key premises in the more interesting argument for each Luminosity thesis.

My response to this argument is to accept that the factive reading of (1)-(3) violates Reasons-Transparency, but to reject Reasons-Transparency anyway. I don't think there is good reason to endorse either Reasons-Luminosity+ or Reasons-Luminosity-. To begin with, let me just state my opposition to the thought that all psychological conditions are transparent. I think Williamson (2000) does a good job of refuting that suggestion and I am not going to say anything more about that matter here. Thus, I reject the first way of motivating Reasons-Transparency.

But what about the more interesting way of motivating Reasons-Transparency? According to the more interesting way of motivating it, the only way to change the rational credentials of one's ϕ -ing would be to engage in reasoning the outcome of which is a change to one's ϕ -ing so that its rational credentials are improved. The change would have to be a change with respect to the reasons for which one is ϕ -ing. The further thought is that this episode of reasoning would have to take a certain shape: it would have to involve the appearance of a second-order knowledgeable judgement about one's ϕ -ing prior to the process of reasoning being undertaken, to the effect that it was, or wasn't, a ϕ -ing for the reason that p. Now if we add to that picture of the sort of reasoning one would have to engage in in order to bring about positive change to the rational credentials of one's ϕ -ing the thought that we are always in a position to affect such changes, so that we are always in a position to engage in such reasoning, we get the result that we're always in a position to know whether or not we're ϕ -ing for the reason that p, when we are ϕ -ing. So what we have here is a conception of what it would take to change the rational credentials of one's ϕ -ing for the better – it would take engaging in an episode of reasoning that involves the manifestation of a piece of knowledge to the effect that one is currently ϕ -ing for the reason that p – added together with the thought that one is always in a position to change the rational credentials of one's ϕ -ing for the better.

I think there is room to challenge either claim. It simply isn't clear what would motivate the thought that one is always in a position to change the rational credentials of one's belief for the better. The thought that we can looks to be an article of faith in the sort of normative luck convincingly attacked by Williams (1976). But I won't pursue that line of response in detail here. Instead, I want to challenge the positive conception of what it would take to change the rational credentials of one's ϕ -ing for the better operated with by the proponent of the argument. Here, there are two claims: (i) that changing the rational credentials of one's ϕ -ing for the better involves engaging in an episode of reasoning and (ii) that the

episode of reasoning would have to involve a second-order judgement that manifests one's knowledge that one is ϕ -ing for a certain reason, or no reason at all.

In response, I wish to accept (i), but to reject (ii). (ii) is objectionable because it is not clear why an episode of entirely first-order reasoning, successfully carried out, should not be capable of ringing positive changes to the rational credentials of one's ϕ -ing. Consider an episode of reasoning which results in the changing of the rational credentials of my belief that Labour will win the by-election. Suppose I reflect on whether I should have that belief. And suppose I held that belief for no reason, prior to my engaging in reflection about it. I might change the status of my belief from irrational to rational just by engaging in first-order reasoning about Labour's chances of victory: reasoning which focuses entirely on the relevant facts about the by-election. There need be no second-order judgement to the effect that I don't currently believe that Labour will win for a reason in play at all. In general, then, one can engage in reasoning which results in positive change to the rational credentials of one's ϕ -ing which does not involve the occurrence of a knowledgeable judgement about whether one's ϕ -ing is currently a ϕ -ing for a reason. (ii) looks false.²³

There is a response to these points available to the proponent of (ii) which involves bringing in the distinction between substantive and structural rationality. The response is that the argument for Reasons-Luminosity under consideration can be read as utilising the notion of substantive rationality or as utilising the notion of structural rationality. But, so the response continues, it is intended to be read as utilising the latter notion. So understood, the argument would be that one is always in a position to change the structural-rational credentials of one's ϕ -ing, and that would indeed involve engaging in reasoning which involves making judgements about the reasons for which one is currently ϕ -ing.

But this response does not succeed. Ringing changes to the structural-rational credentials of one's ϕ -ing either consists in engaging in reasoning which involves second-order judgement or it doesn't. If it does, the second-order judgement would take the shape not of a judgement about the reasons for which one is ϕ -ing, if any, but a judgement to the effect that one is in certain mental states which make it structurally rational for one to ϕ . One could then reason from such second-order propositions to a judgement about whether one ought, in the structural sense, to ϕ and thus to changing the structural-rational credentials of one's ϕ -ing. But if that's the way the reasoning which results in positive change to the structural-rational credentials of one's ϕ -ing works, then one can at best conclude, in conjunction with the thought that we are always in a position to change for the better the structural-rational credentials of our ϕ -ing, that we are always in a position to know what rationality-making mental states we're in. But whatever the truth of that claim, that is not what Reasons-Luminosity says.

On the other hand, ensuring that one is ϕ -ing in a structurally rational way might involve engaging only in first-order reasoning. This is indeed how ϕ -ing in a structurally

²³Compare Williamson (2000:180, 192).

rational way works, I think: roughly, one counts as coming to ϕ in a structurally rational way just by engaging in reasoning which one cannot distinguish from an episode of reasoning which would be coming to ϕ in response to a genuine reason. But, as we've seen, the latter sorts of episodes need involve the invocation of no second-order judgement. We can conclude, then, that explicitly operating with the notion of structural rationality, instead of which the notion of substantive rationality, is of no help to the proponent of the current argument for Reasons-Transparency.

The upshot is that there is nothing to be said in favour of Reasons-Transparency. The mundane way of motivating it is objectionable. Moreover, the more interesting way of motivating it relies on a premise to the effect that we're always in a position to change the rational credentials of our ϕ -ings and the premise that changing the rational credentials of our ϕ -ings is necessarily a matter of engaging in reasoning which is partly second-order. The first premise is undermotivated and the second looks false. The proponent of the factive reading of (1)-(3) can thus reject the Transparency Thesis as undermotivated in turn.

3.7 The Motivating Reasons View Rejected

I've defended the factive reading of (1)-(3) against four arguments to the contrary. ϕ -ing for a reason is always to be thought of as ϕ -ing in response to a normative reason. What effect does this have on the issue of whether the Motivating Reasons View is correct?

Well, it looks like the Motivating Reasons View is rendered false by our result, because it looks like our result renders the Motivating Reasons Postulate false. It renders the Postulate false because according to the latter, when the agent ϕ s for a reason, so that (1)-(3) are true of them, the sense in which the agent is motivated by a *reason* is not that they are motivated by a normative reason. Rather, they are motivated by an item which has a special status of reason-hood individuated by the rationalising role it plays, where it plays that rationalising role across both the good case and the bad case. But the argument here has established that (1)-(3) involve the normative sense of reason-hood. So the postulate is false: there are no motivating reasons.

It's important to stress that the claim that there are no motivating reasons sounds needlessly paradoxical. That's because there is an ordinary phenomenon tracked by talk of the reasons which motivate. The ordinary phenomenon is what's captured by (1)-(3). When I say that there are no motivating reasons it might seem as if, absurdly, I'm denying the existence of that phenomenon. But that's not what I'm doing. Rather, I'm denying that a certain interpretation of that phenomenon is correct – the interpretation embodied by the Motivating Reasons Postulate. That's how the slogan that there are no such things as motivating reasons should be understood.

But the proponent of the Motivating Reasons Postulate has a way of responding to

what I've said here. They might try to say that even if the factive reading of (1)-(3) is true, it doesn't follow that the Motivating Reasons Postulate, or something true to the spirit of it anyway, is correct. That's because the Motivating Reasons Postulate effectively tells us that the phenomenon of ϕ -ing for a reason is a phenomenon which consists in the obtaining of a state of affairs which grounds the truth of a good case/bad case neutral rationalising explanation, where the explanation in question takes as its *explanans*, either wholly or partly, an item which counts as a reason in a sense different from the sense in which normative reasons are reasons. (1)-(3) do capture a phenomenon worth calling ' ϕ -ing for a reason', and it is indeed true that that phenomenon shouldn't be modelled in the way described by the Postulate. To the extent that the Postulate was intended to offer us a theory of the phenomenon captured by (1)-(3), then, it is false. But given what's been said here, (4)-(6) read non-factively also capture a phenomenon worth calling ' ϕ -ing for a reason', and that phenomenon, which is distinct from that captured by (1)-(3), is apt for an analysis in terms of the Motivating Reasons Postulate. So the Postulate can live on, in spite of what's been said here.

The problem with this is that the phenomenon captured by (4)-(6) read non-factively do not apply the concept of a reason to anything. So it cannot be that the Postulate is true of the phenomenon captured by (4)-(6) because the Postulate requires that the phenomenon in question is one which involves something to which we're willing to apply a concept of a reason. There is no way to resurrect the Motivating Reasons Postulate: there really are no such things as motivating reasons.

3.8 Re-Drawing the Map of the Terrain

The Motivating Reasons View has been refuted. Where does this leave the issues flagged up in the last chapter? In particular: where does it leave the Dialectic of Psychologism, which seems premised on the truth of the Motivating Reasons View, and where does it leave the second component of the Motivating Reasons View – the Primacy of Motivating Reasons? In this section I'm going to argue that each of those issues is still live even though the Motivating Reasons View has been refuted. We'll just have to be careful to reformulate them in the light of the rejection of the Motivating Reasons Postulate.

3.8.1 The Dialectic of Psychologism

There are three parties to the Dialectic of Psychologism: Davidson, Dancy, and the Neo-Davidsonians. All players in the debate are proponents of the Motivating Reasons View, and so would deny the factive reading of (1)-(3). They would all agree that (1)-(3), (4)-(6) and (7) are true across both good cases and bad, and that they should be interpreted as applying a concept of a reason distinct from the concept of a normative reason to the

relevant item and as saying that S is ϕ -ing in response to that item. They would take each of those sentences to report the sort of neutral rationalising explanation that they think is the only type of rationalising explanation, although there would be disagreement about the nature of the rationalising explanation in question, as we've seen.

Now, there are two issues at the heart of the Dialectic, as standardly conceived. First, there's the issue about whether to think of motivating reasons as psychological items. Second, there's the issue of whether to think of rationalisers as motivating reasons. The dialectic gets under way because it looks implausible to identify motivating reasons with psychological items instead of with what the subject takes to be a normative reason. But it looks implausible also to identify rationalisers with motivating reasons conceived of in the non-psychologistic way, for then we'd have to deny that rationalising explanation is factive. The three players in the debate stake out different territories in this dialectical terrain.

However, I take it it's possible for there to be a debate about the two issues that's independent of the truth of the Motivating Reasons View. The issues can be reformulated in a way that doesn't require us to endorse the existence of motivating reasons in the special sense without significant loss. Take the first issue first. Stripped of the usual presumption in favour of motivating reasons, the issue boils down to this: are the reasons which motivate the agent to ϕ to be identified with psychological items or not? And the second issue boils down to this: must rationalisers be identified with the reasons which motivate the agent to ϕ ? The notion of a reason which motivates in play here is the ordinary notion expressed in sentences (1)-(3), not the special notion of a motivating reason. The players in the Psychologism Dialectic make the mistake of identifying that ordinary notion with the special notion of a motivating reason, and so they end up casting the debate, and their positions with respect to it, in the terms they do. But a neutral casting of the debate is possible, and that's just what I've supplied.

It might be that re-cast, the issues here are more easily solvable than they were when cast in terms of the special notion of a motivating reason. But the point for the moment is just that they can be re-cast, so that the debate is still live. I attempt to settle the re-cast first issue in the next chapter and I attempt to make some initial headway of resolving the re-cast second issue there too.

3.8.2 The Primacy of Motivating Reasons

The Primacy of Motivating Reasons says that we should offer a Reductive account of what it is for an agent to ϕ in response to a normative reason, where that Reductive account is cashed out in terms of the special notion of a motivating reason, as defined by the Motivating Reasons Postulate.

It should be easy to see how to abstract the central issue here from the Motivating

Reasons Postulate: it is possible to subscribe to a Reductionist account of the good case whilst denying the Postulate. That would amount to saying that the success phenomenon of ϕ -ing in response to the normative reason that p consists in ϕ -ing in a way that's subject to the sort of neutral rationalising explanation paradigmatically reported by psychologicistic rationalising statements, plus further factors present in the good case which are, together with the explanatory condition, sufficient for one to count as ϕ -ing in response to a reason.

The Primacy of Motivating Reasons makes the mistake of taking it that rationalising explanations are always to be written in terms of the special notion of a motivating reason, and thus makes the mistake of thinking that a Reductive account of the good case must always take the shape of an account which reduces ϕ -ing in response to a normative reason to ϕ -ing for a motivating reason, plus whatever further factors need to be added in order to generate success. But Reductionism need only take this form if there are such things as motivating reasons, and it's been established that that's an empty category. From now on, I'll be understanding Reductionism in the neutral way just described.

Again, one might think that the material presented in this chapter puts us in a better position to refute the Reductionist account of the good case than we were when the Motivating Reasons View was a live option. Whatever the truth of that idea, all that's important to note for now is that offering a Reductive account of the good case needn't take the form of accounting for the good case in terms of ϕ -ing for a motivating reason plus extra factors. It can take the more neutral form of ϕ -ing in a way that's subject to a neutral type of rationalising explanation, which is the only type of rationalising explanation there is, plus the relevant extra factors.

3.9 A Merely Linguistic Dispute?

In this chapter I've been engaging in a debate that has been advertised as a debate about the meaning of certain types of sentence: (1)-(3). I've argued that we need to read them as applying the concept of a normative reason, so that they report the condition of ϕ -ing in response to a normative reason and are hence only true in good cases. But why is such a linguistic debate interesting from the philosophical point of view? We might have learned something about the English language: that certain types of sentences belonging to it are true only in certain sorts of cases, and that the word 'reason' doesn't have a distinctive meaning when it comes to the notion of being motivated by a reason, but why should those linguistic results be of any interest?

One immediate point to make is that the linguistic results here refute the Motivating Reasons Postulate, which is itself a significant result.

A further quite weak point to make at the outset is of an *ad hominem* character. Philosophers like Dancy, Turri, Setiya, Schroeder, Comesaña and McGrath, and many others who haven't been explicitly discussed here, clearly ascribe philosophical importance to the idea

that the agent ϕ s for a reason even in the bad case. That claim is put to work in arguments by some of those philosophers, whereas others go to significant lengths in order to argue for it. The philosophers in question clearly ascribe importance to it, then, for whatever reason. The present chapter refutes the claim that agent's ϕ for reasons in bad cases. That result will trouble the philosophers in question and that in itself carries some significance.

But the results here are not merely linguistic anyway. For the results here tell us something about the *metaphysics* of ϕ -ing for reasons, and they tell us something about the *conceptual* repertoire we employ when we think of agents as ϕ -ing for reasons, as I now wish to detail.

First, the metaphysical consequences of what has been established here. I take it that establishing that (1)-(3) are true only in the good case tells us something about *what it is* to ϕ for a reason. It tells us that ϕ -ing for a reason consists not in standing in some relation to a rationalising item that has the status of a reason and which is present, playing the relevant role, even in the bad case. Rather, ϕ -ing for a reason consists in standing in some relation – the ‘responding to’ relation – to an item which favours one's ϕ -ing.²⁴

Second, there are conceptual consequences of what has been established here. I take it that establishing that (1)-(3) are true only in the good case tells us something about what the concept of a reason in play in (1)-(3) amounts to. It tells us that that concept is the concept of a normative reason. An upshot of this is that at most, we have only two concepts of a reason: the concept of a normative reason and the concept of an explanatory reason. We apply the former concept in cases where we're thinking of a fact as having a certain sort of normative dimension, and in contexts in which we're thinking of the agent as ϕ -ing for a reason. We apply the latter whenever we endorse an explanation why something is the case. The two concepts might well overlap: it might be that a fact could be both a normative reason and an explanatory reason and indeed later on I'll be arguing that in addition to the sort of neutral rationalising explanations reported paradigmatically by psychologistic rationalising statements there is a type of rationalising explanation which takes as *explanantia* normative reasons. Normative reasons can be the *explanantia* of such rationalising explanations precisely because of the normative dimension they display. So later on I'll be arguing not just that normative reasons can also be explanatory reasons, but also that they get to play the second role by dint of playing the first.²⁵

So I have not merely been engaging in a linguistic dispute in this chapter. Another way in which the results of this chapter are significant is that they provide us with a basis for a response to an influential argument in favour of Psychologism, as we shall see in the next chapter. Moreover, they help us undermine the conception of rationalising explanation at the heart of the Reductionist's position, as we shall see in Chapter Five.

²⁴Indeed, I've sometimes switched from a linguistic way of talking about the issues to a metaphysical way of talking about them.

²⁵This is a position defended by Raz (2011). See also Alvarez (2010) and Roessler (2014).

Chapter 4

The Dialectic of Psychologism

Consider the following passage from Michael Smith's *The Humean Theory of Motivation*:

... the claim that is, as I understand it, constitutive of the Humean theory – is the claim that motivation has its *source* in the presence of a relevant desire and means-end belief. This claim finds more formal expression in the following principle:

P1. R at t constitutes a motivating reason for agent A to ϕ iff there is some ψ such that R at t consists of a desire of A to ψ and a belief that were he to ϕ he would ψ .

(Smith, 1987:36)

Here Smith identifies the reasons which motivate the subject to act with states of believing and desiring belonging to the subject: the desire specifying the goal the agent pursues in acting, the belief the information that so acting is a way of achieving that goal. These states are the subject's motivating reasons for action, and they are distinguished from the agent's normative reasons for action.¹ Smith is a proponent of Psychologism about the agent's reasons for action. He is also a proponent of the Motivating Reasons View, so that his preferred version of Psychologism is written in terms which presuppose the existence of motivating reasons in the special sense.

There are no such things as motivating reasons in the special sense, as we have seen. So Smith's view is already in need of some modification. However, even if we reject the Motivating Reasons View we are still left with the following question: should we identify the reasons for which the subject ϕ s with psychological items possessed by the subject? We're ignoring the contribution made by any relevant conative state in this area and focusing entirely on the belief the subject has, the content of which is what gets taken to be a normative reason to ϕ by the subject. So the question becomes: should we identify

¹Smith (1987:§.2)

the reasons for which the subject ϕ s with states or facts of believing that p, where p is what the subject takes to be a normative reason for them to ϕ ? This chapter argues for a negative answer to that question and instead suggests, following Dancy and a good deal of other contemporary philosophers, that the reasons which motivate the agent to ϕ are what the subject takes to be the reasons in favour of their ϕ -ing.

However, that answer to the first question complicates our assessment of a second question: should we identify the reasons which motivate the agent to ϕ with the *explanans* of the rationalising explanation to which the agent's ϕ -ing is subject, when they ϕ for a reason? If we say that the reason for which the subject ϕ s is just that p, then won't we have to deny that explanation why is factive, as Dancy does?

All of this is just to introduce, in a different way, the Dialectic of Psychologism, which should already be familiar. There are two issues to be addressed here:

The Psychologism Question. Are the reasons for which the subject ϕ s to be identified with psychological items belonging to the subject?

The Explanatory Question. When the agent ϕ s for a reason, is the *explanans* of the rationalising explanation to which their ϕ -ing is subject to be identified with the reason for which they ϕ ?

To repeat, these issues are to be interpreted here without a presumption in favour of the Motivating Reasons View, so that the notion of a reason which motivates the subject to ϕ at work in the questions above is the ordinary notion utilised by (1)-(3), not the special notion of a motivating reason, which has been shown to be empty.

There are two preliminary points to be made. First, the phenomenon we're focusing on is the phenomenon of ϕ -ing for the reason that p. That is the phenomenon paradigmatically reported using sentences (1)-(3). Putting aside whether the factive reading of (1)-(3) is true for a moment, isn't the Psychologism Question simply settled in favour of Anti-Psychologism anyway, given that sentences (1)-(3) quite clearly identify what the agent takes to be a normative reason – namely p – with the reason for which the agent ϕ s? And if that's so, why do we need to give that question substantive treatment in this chapter? Some thought like this is what motivates Setiya to suggest that:

The doctrine of 'psychologism', according to which the reasons for which we act are always states of, or facts about, our own psychology, is hopeless.
(Setiya, 2011:132)

The response to this is simply that the proponent of Psychologism does indeed have ways of dealing with the worry, as we shall see in more detail later on. Thus, the present thought that the debate over Psychologism can simply be finessed in favour of Anti-Psychologism in the way described is, it will turn out, untrue.

I will proceed in this chapter as follows. §§4.1-4.2 take up the Psychologism Question. §4.1 presents and defends two arguments in favour of Anti-Psychologism. §4.2 identifies and rejects two arguments in favour of Psychologism, which I take to be the strongest available. Finally, §4.3 makes a start on the Explanatory Question. The Explanatory Question cannot be completely answered prior to the results of Chapter Five, in which it is argued that there are *two* types of rationalising explanation: a neutral type the *explanans* of which is the fact that the subject believes that p and an essentially successful type, the *explanans* of which is the fact that p, *qua* normative reason. The burden of §4.3 is to merely unpack my positive answer to the Explanatory Question, how the conclusion of the next chapter delivers it, and to relate it to the positions on the Explanatory Question already presented. This sets things up for a solution to the Explanatory Question to come as a corollary of the argument of the next chapter.

4.1 Two Arguments for Anti-Psychologism Defended

This section presents and defends two arguments for Anti-Psychologism. §4.1.1 focuses on a linguistic argument and §4.1.2 focuses on an argument which appeals to the idea that Psychologism has some counter-intuitive results. Before I begin, a preliminary point is in order. I distinguished in Chapter Two between a state version of Psychologism and a fact version of Psychologism. This section provides arguments against Psychologism *in general*. Each argument to be considered is effective against both versions of Psychologism.

4.1.1 The Linguistic Argument

The first argument I focus on is *The Linguistic Argument*.² The claim to be established is that when the agent ϕ s for a reason, the reason for which they ϕ is that p, and not their believing that p or that they believe that p. Presented slightly differently: the condition of being motivated to ϕ by a reason – the condition picked out by (1)-(3) – is a condition which involves an item that has the status of a reason, and the claim to be established is that the item in question is p, not the agent's believing that p or that they believe that p.

The linguistic argument focuses on (4)-(6):

- (4) S's reason for ϕ -ing is that p
- (5) S's basis for ϕ -ing is that p
- (6) S's ground for ϕ -ing is that p

The proponent of the Linguistic Argument says that the correct semantics for (4)-(6) implies that Anti-Psychologism is true.

²Proponents of the Linguistic Argument include: Miller (2008) and Alvarez (2010:131).

In order to enable us to get a grip on the Argument, let's focus on the following instance of (4):

(4a) My reason for believing that Labour will win the by-election is that the exit-poll predicts a Labour victory

The proponent of the Linguistic Argument makes the following claims about (4a): (i) the copula 'is' which appears in (4a) is the 'is' of identity, and (ii) the that-clause which appears in it denotes what the subject believes. So the proponent of the Linguistic Argument says that we should think of (4a) as *identifying* my reason with what I believe: that the exit-poll predicts a Labour victory. More generally, the proponent of the Linguistic Argument says that any sentence of form (4)-(6) is in the business of identifying the subject's reason with what they believe about the situation.

It is not difficult to see how this supposedly natural semantics for (4)-(6) delivers us the result that an Anti-Psychologism about what gets characterised by (4)-(6) as the subject's reason/basis/ground is true. For on the natural semantics the subject's reason is identified by those sentences with the referent of the relevant that-clause. And the referent of the relevant that-clause is what the subject takes to be a normative reason, it is not the subject's state of believing that p or the fact that they believe it.

If we were working with the assumption that (4)-(6) always mean the same as (1)-(3) then we could readily conclude that Anti-Psychologism is correct. For Anti-Psychologism is a thesis concerning what gets picked out by (1)-(3), and if (4)-(6) just mean the same as (1)-(3) then, given that (4)-(6) identify the subject's reasons with what they believe, as opposed to their believing it or that they believe it, we'd get the result that (1)-(3) do so too.

But we are precisely not operating with the assumption that (4)-(6) must always mean the same as (1)-(3). How, then, do the linguistic considerations which yield an Anti-Psychologism about what gets characterised as the subject's reason by (4)-(6) deliver us Anti-Psychologism about ϕ -ing for reasons – the condition denoted by (1)-(3)?³

(4)-(6) admit of a factive reading, on which they do mean the same as (1)-(3) and a non-factive reading, on which they don't. On the first, the phrase 'the subject's reason' is intended to characterise the item in question as a reason. On the second, the phrase serves to characterise the item only as something that counts as a normative reason from the subject's own point of view, and hence doesn't apply a concept of a reason to the item in question.

Read either way, the semantics described has initial plausibility. The only difference would be that read in the factive way, the semantics generates the result there is something

³ Alvarez and Miller are insensitive to this question because they assume that (4)-(6) must always mean the same as (1)-(3). What follows, then, should be advertised as a developed version of the argument they offer. They are not, strictly speaking, proponents of precisely the argument I state here.

that is designated as a reason by (4)-(6), and that item is identified by (4)-(6) with the referent of the relevant that-clause. On the non-factive reading, what gets identified with the referent of the relevant that-clause is not something which also gets characterised as a reason by (4)-(6) but something that gets characterised as an apparent normative reason.

On the factive reading, (4)-(6) have the same meaning as (1)-(3). The semantics described is plausible for the factive reading and it establishes an Anti-Psychologism about the condition reported by (4)-(6) on the factive reading. So we get the result that Anti-Psychologism concerning the condition reported by (1)-(3) is true too, which just is our desired conclusion.

It's worth noting that there is a more direct way of utilising linguistic considerations in order to prove Anti-Psychologism. Let's focus on (1b)-(3b):

(1b) The reason for which S ϕ ed/is ϕ -ing is that p

(2b) The basis on which S ϕ ed/is ϕ -ing is that p

(3b) The grounds on which S ϕ ed/is ϕ -ing is that p

The emphasised copula in each of (1b)-(3b) can naturally be read as an 'is' of identity. Thus, (1b)-(3b) seem to identify the reason/basis/grounds on which S has ϕ ed or is ϕ -ing with what gets denoted by the relevant that-clause: what the subject believes, and not their believing of it, or the fact that they believe it. Since (1b)-(3b) quite clearly mean the same as their (1a)-(3a) counterparts and hence denote the condition of ϕ -ing for a reason, it follows directly that Psychologism is false and Anti-Psychologism true. In short, the same style of argument which takes us from (4)-(6) to Anti-Psychologism can be applied to (1b)-(3b) to get us that same result more directly.

The phrase *the Linguistic Argument* will be used to denote the argument which takes us from the supposedly natural semantics for (4)-(6), to the truth of Anti-Psychologism, via the thought that the semantics is plausible for (4)-(6) on the factive reading. The more direct alternative to the Linguistic Argument will feature in the discussion below but because much of the literature in this area focuses on (4)-(6) instead of on (1b)-(3b),⁴ I will continue to focus for the most part on the former set of sentences and the argument for Anti-Psychologism associated with it.

That concludes my presentation of the Linguistic Argument. I now want to describe and respond to a problem with it. The basic problem is that there is an alternative semantics for (4)-(6) available which coheres with Psychologism. I call this alternative semantics *the Pryor semantics*, because it was first suggested by Pryor (2007). I'll refer to the supposedly natural semantics as *the orthodox semantics*.

⁴Including, as we will see momentarily, work which has been done by way of refuting the Linguistic Argument.

According to the Pryor semantics, we should accept (ii),⁵ but reject (i) in favour of the claim that the copula which appears in (4a) functions to specify the content of a mental state picked out by the singular term ‘my reason’, as it appears in (4a). On this view, we shouldn’t think of (4a) as telling us that the agent’s reason is identical to what gets designated by the relevant that-clause. Rather, we should think of the phrase ‘my reason’ as it appears in (4a) as denoting a mental states of the subject. And we should think of the copula as functioning to specify the content of the mental state in question, by appeal to what gets denoted by the relevant that-clause. On the Pryor semantics, then, (4a) doesn’t identify the agent’s reason with anything at all, but rather picks out a certain mental state which is itself to be thought of as the agent’s reason and offers us a specification of its content. More generally, Pryor suggests that (4)-(6) are not in the business of identifying the agent’s reasons/bases/grounds with what gets denoted by the relevant that clause, but rather serve to specify the content of the mental state picked out by ‘S’s reason/basis/ground’ by appeal to the referent of the that-clause.⁶

An analogy which Pryor himself draws with propositional attitude ascriptions might help to bring the Pryor semantics into clearer focus. Consider the following pair:

(8a) S’s belief is that p

(8b) S’s belief is long-standing

According to a familiar view about propositional attitude ascriptions, such ascriptions are ambiguous between a state and a content reading. On a state reading, the propositional attitude ascriptions tells us something about the subject’s state of believing. On the content reading, the ascriptions tells us something about what is believed: the content of the subject’s belief.

That standard view about propositional attitude ascriptions is supported by reflection on (8a) and (8b). For let’s suppose that ‘S’s belief’ always refers to S’s state of believing. Since it’s natural to think that the ‘is’ which appears in (8a) is the ‘is’ of identity, we’d get the result that (8a) is at least false if not unintelligible because it identifies S’s state of belief with what gets denoted by the that-clause which appears in it. So we must instead read ‘S’s belief’ as it appears in (8a) as denoting the content of S’s belief. That way we can read the whole sentence as identifying the content of S’s belief with the referent of the relevant that-clause. But we shouldn’t think of ‘S’s belief’ as it appears in (8b) as referring to the content of S’s belief. For the content of S’s belief is not itself long-standing. What’s

⁵ Although Pryor does in fact raise several objections to the claim that that-clauses are referring devices.

⁶ Strictly speaking, Pryor distinguishes between the familiar ‘is’ of predication from what he calls the ‘is’ of specification. He argues against identifying the copula which appears in (4)-(6) with the ‘is’ of predication (Pryor, 2007:§.6) and instead suggests that we should identify it with the ‘is’ of specification (*ibid.*:§.7). I run roughshod over this distinction because it isn’t necessary to go into it for my purposes.

being referred to and qualified as long-standing is, rather, S's state of believing. Thus, we get the result that 'S's belief' has a state and a content reading.⁷

For reasons that need not concern us, Pryor argues that we should reject the orthodox suggestion that 'S's belief' is ambiguous between a state and a content reading. He suggests instead that we should think of 'S's belief' as always referring to S's state of believing and we should reject the idea that the copula in (8a) is an 'is' of identity. Instead, Pryor suggests, we should read the copula as functioning to specify the content of the belief by appeal to what gets denoted by the relevant that-clause. (8a) says that S's belief has the content: p. Both (8a) and (8b) talk about S's state of believing.

What goes for (8a) goes for (4a) and for (4)-(6) more generally, according to Pryor. Just as we should think of 'S's belief' in (8a) as referring to S's state of believing and then read the whole sentence as specifying the content of the belief, so we should think of 'my reason' in (4a) as referring to a mental state and read (4a) as specifying the content of the state in question by appeal to what gets denoted by the relevant that-clause.

It's easy to see how the Pryor semantics would, if true, undermine the Linguistic Argument. Were the Pryor semantics true, the orthodox semantics for (4)-(6) would be false. So the starting point for the Linguistic Argument would be undermined. This objection to the linguistic argument however, is not decisive, I have only tried to show that the argument as it stands is inconclusive. The thought is that *as long as there is no argument in favour of the orthodox semantics or against the Pryor semantics offered*, the Linguistic Argument is dialectically ineffective.

Moreover, the Pryor semantics, if true, rules out the more direct linguistic argument for Anti-Psychologism which appeals to (1b)-(3b). For it is possible to construct an analogue of the Pryor semantics for (1b)-(3b). The analogue semantics would say that the emphasised copula in (1b)-(3b) is not an 'is' of identity but an 'is' of specification. What the referent of the relevant that-clause denotes is the content of the mental state which gets referred to by the descriptive phrases 'the reason for which', 'the basis on which', and 'the grounds on which' in (1b), (2b), and (3b) respectively. The ensuing dialectical situation would then be the same for the proponent of the idea that the correct semantics for (1b)-(3b) deliver us the truth of Anti-Psychologism.

So the possible truth of the Pryor semantics undermines the Linguistic Argument and its more direct cousin. I want to reject the Pryor semantics in favour of the orthodox semantics, thus rescuing the Linguistic Argument and its more direct cousin from the dialectical bind they find themselves in.

Before I do that, however, there is a point worth drawing attention to. I noted in the introduction to this chapter that Psychologism looks to be a doctrine that's obviously false, given that (1)-(3) seem clearly to identify p with the reason for which the subject ϕ s. But we're now in a position to see how the proponent of Psychologism might accept (1)-(3)

⁷The same goes for analogous noun phrases associated with other types of propositional attitude.

consistently with their view that the reason for which the subject ϕ s is their belief, or the fact that they have the relevant belief. For the proponent of Psychologism can and should accept the idea that whenever we have a sentence which concatenates a copula with a that-clause so that it seems as if something that's being designated as a reason is being identified with the referent of the relevant that-clause, what is really going on is that the copula is functioning to specify the content of a mental state of believing being picked out by whatever phrase is doing the work of characterising the relevant item as a reason.⁸ The proponent of Psychologism can and should adopt the Pryor semantics for (4)-(6) and an analogue of the Pryor semantics for any other sentence types – including (1b)-(3b) – that would look to cause them trouble. This is enough to fend off Setiya's suggestion that the Question of Psychologism can simply be finessed in favour of Anti-Psychologism.

The Pryor semantics for (4)-(6) is, however, implausible. I now want to raise two problems for it. The first is an objection, the second is an attempt to shift the burden of proof.

The first problem I want to raise is that if the Pryor Semantics is correct, then we should expect to be able to make sense of sentences which predicate properties to what gets denoted by the expression 'S's reason', which appears in sentences of type (4), that are properties of propositional attitudes the subject has. The issue is that such sentences sound very odd, and appear to constitute category errors. Consider the following sentences:

(9a) S's reason for ϕ -ing has persisted throughout S's life

(9b) I caused S's reason for ϕ by telling her that p

(9c) S's reason for ϕ -ing is justified

(9d) S's reason for ϕ -ing was hastily formed

If the Pryor semantics is right, then we should be able to make ready sense of (9a)–(9d). But, as I say, the sentences are odd to say the least. Suppose, for example, that Tim's reason for riding his bike to work is that it will cut down on petrol costs. Does it sound correct to say, as the Pryor semantics seems committed to thinking it is, that Tim's reason has persisted throughout his life?

⁸It's worth noting at this stage that the Pryor semantics, were it true, wouldn't cohere well with fact Psychologism, only state Psychologism. For the copula is said, by the proponent of the Pryor semantics, to specify the content of a state, not to characterise the character of a certain fact. Indeed, it's not clear whether there is a type of copula which performs a role with respect to facts as Pryor's specificational copula does with respect to states of mind. Thus, linguistic considerations might still be thought to count against fact Psychologism, for the only options with respect to the semantics of (4)-(6) – the orthodox semantics and the Pryor semantics – get us the result that fact Psychologism is false. But that doesn't serve to undermine the main point being made here, which is that the availability of the Pryor semantics renders the Linguistic Argument inconclusive. For that argument purports to rule out Psychologism *in general*, and thus prove Anti-Psychologism true.

Now, it should be noted that the claim here isn't that (9a)-(9d) sound odd no matter what the corresponding instance of (4)-(6) is. If S's reason for ϕ -ing is that they believe that p, or that they desire to ψ , for example, then the relevant instance of (9a)-(9d) might be readily intelligible – 'S reason for ϕ -ing has persisted throughout S's life' makes ready sense if the corresponding instance of (4) is 'S's reason for ϕ -ing is that they desire to ψ ', for example. The present point is just that (9a)-(9d) should always be readily intelligible if the Pryor semantics is true, no matter what the corresponding instance of (4)-(6) is. But that isn't so, for when S's reason for ϕ -ing is some state of the world, as it is in the example I selected above, (9a)-(9d) will come out as counter-intuitive.

Moreover, the objection here can be extended to the analogue of the Pryor semantics for (1b)-(3b), thus rescuing the orthodox semantics for those sentence types and rescuing, in turn, the more direct linguistic argument for Anti-Psychologism. Consider the following:

- (10a) The reason for which S ϕ s has persisted throughout S's life
- (10b) I caused the reason for which S ϕ s by telling her that p
- (10c) The reason for which S ϕ s is justified
- (10d) The reason for which S ϕ ed was hastily formed

If the analogue of the Pryor semantics for (1b)-(3b) were true, then we should be able to make ready sense of (10a)-(10d), at least when the corresponding instances of (4)-(6) concern states of the world. But, again, such sentences sound quite odd indeed.

There is a second problem for the Pryor semantics which I now want to discuss, the point of which is to shift the onus of proof onto the proponent of it. It concerns what sense the proponent of the Pryor semantics can make of the idea that it is possible for agent's to ϕ on the basis of *good* or *the right* reasons, statuses which I take to amount to the same thing.

What I want to suggest, first of all, is that the proponent of the Pryor semantics is committed to thinking that when we say of an agent that they have ϕ ed or are ϕ -ing for good reasons, what we are doing is predicating the mental state of believing which the proponent of the view is committed to thinking of as the agent's reason. The phrase 'good reason' in other words, functions to predicate some property of the mental state of believing picked out by the phrases 'S's reason/basis/grounds for ϕ -ing' as it appears in sentences (4)-(6), according to the Pryor Semantics.

To see this, consider the sentences:

- (11) Arthur's reason for visiting the zoo is a good one

The proponent of the orthodox semantics is committed to saying, about (11), that it predicates the property of being a good reason to what gets denoted by the that-clause

which would appear in the corresponding instance of (4). For they think the referent of ‘Arthur’s reason’ is *identified* with the referent of the that-clause by the corresponding instance of (4), and (11) predicates the property of being a good reason to whatever the referent is of the relevant that-clause. So (11) ascribes the property of being a good reason to what the subject believes.

The proponent of the orthodox semantics can thus say that when we ascribe the property of being a good reason to the subject’s reason using a sentence like (11), the property we have in mind is simply the property of being a normative reason – the property of being a reason in favour of the subject’s ϕ -ing. That’s because what the agent believes is the sort of thing that could be a normative reason – it is the sort of thing that could be a fact which favours ϕ -ing. So what the proponent of the orthodox semantics could say about (11) is that it says that Arthur is acting on the basis of a genuine reason in favour of his action, although it does not tell us what that reason is. They can say this, moreover, whether the phrase ‘Arthur’s reason’ is read factively or non-factively, although if it were read factively then (11) would be trivial.

The proponent of the Pryor semantics, by contrast, is committed to saying that the referent of the phrase ‘S’s reason’ is a mental state of believing the agent is in – a mental state which partly constitutes the subject taking a certain factor to be a normative reason for their ϕ -ing. It is that mental item which is the agent’s reason for ϕ -ing. So what is being predicated with the property of being a good reason by (11) is a mental state of believing the subject is in, not what the agent believes.

The upshot of this is that the proponent of the Pryor semantics is committed to saying something particularly contentious about what property we have in mind when we predicate the property of being a good reason using a sentence like (11). What they are committed to saying is that the property being predicated cannot be the property of being a normative reason, for the agent’s belief is not a reason in favour of their ϕ -ing.

Why is it problematic to deny that the property of being a good reason ascribed by sentences like (11) is not the property of being a normative reason? To see why, consider the following inference:

Good Reasons Inference

- (I) S’s reason for ϕ -ing is that p
- (II) S’s reason for ϕ -ing is a good reason

- (C) P is a good reason for S to ϕ

The Goods Reasons Inference is valid, on the face of it. One way of explaining why it’s valid is that (I) *identifies* S’s reason with p and (II) ascribes the property of being a

normative reason to whatever S's reason is, thus giving us the result that p has the property of being a normative reason for S to ϕ , which is what (C) says. That account of the validity of the inference is open only to the proponent of the orthodox semantics, of course.

Thus, the proponent of the Pryor semantics owes us a fresh account of the validity of the inference which coheres with their favoured semantics for sentences (4)-(6) and (11). But I don't want to press the point that there is no such account available. What I want to suggest is simply that at first glance, the Good Reasons Inference concludes that p is a normative reason for S to ϕ . After all, that is what sentences of type (C) are usually used to assert, and we'd need an argument to convince us that in the context generated by the Good Reasons Inference, things are radically different. And, moreover, at first glance the argument utilises a single sense of good reasonhood. On the face of it, that is, there is no transition from one use of the phrase 'good reason' in (II) to a second, quite different use of the phrase by (C). On the face of it, the phrase 'good reason' is used univocally.

This is not, of course, a decisive point against the Pryor semantics. But it does serve to shift the onus of proof onto the proponent of the Pryor semantics. Unlike the orthodox semantics, the Pryor Semantics is saddled with the thought that when we utter sentences like (11) we are ascribing a property using the phrase 'good reason' which is different from the property of being a normative reason, and they are thus committed to thinking that (C) isn't ascribing the property of a normative reason to p or else that it is but that the Good Reasons Inference doesn't use the phrase 'good reason' univocally. On the face of it, neither of those claims is true. Thus, the proponent of the Pryor Semantics will need to present us with an argument for their favoured semantics. In the absence of such an argument, the Pryor Semantics should be rejected in favour of the Orthodox Semantics.

I've been focusing on the Linguistic Argument and thus on sentences (4)-(6), but an analogous point holds with respect to the more direct cousin on the Linguistic Argument which focuses instead on (1b)-(3b). The analogue of the Pryor semantics with respect to those sentences is likewise committed to the thought that 'the reason/basis/grounds for which/on which S ϕ s', as those expressions appear in (1b)-(3b), function to pick out mental states of the subject and hence when we say of the item picked out by those phrases that it is a *good reason* what we're doing is ascribing a certain property to the relevant mental state. An analogue of the Good Reasons Inferences involving (1b)-(3b) would then prove that the onus is on the non-orthodox treatment of (1b)-(3b).

In conclusion then, merely pointing out that there is an alternative semantics for (4)-(6) available which undermines the Linguistic Argument is no good because the alternative semantics is one which will need to be motivated, if it's to be acceptable. The orthodox semantics should be our default. Moreover, the alternative semantics is independently objectionable anyway. These points apply in equal measure to the analogue of the Pryor semantics for (1b)-(3b) which is supposed to undermine the more direct cousin of the Linguistic Argument too.

4.1.2 The Counter-Intuitiveness Argument

In this section I develop and defend an argument for Anti-Psychologism promoted by Dancy (2000). Let's begin with a statement of the argument. Dancy markets the argument as an argument against the fact version of Psychologism.⁹ Consider once again statements (1)-(3):

- (1a) S ϕ ed/is ϕ -ing for the reason that p
- (1b) The reason for which S ϕ ed/is ϕ -ing is that p
- (2a) S ϕ ed/is ϕ -ing on the basis of p
- (2b) The basis on which S ϕ ed/is ϕ -ing is that p
- (3a) S ϕ ed/is ϕ -ing on the grounds that p
- (3b) The grounds on which S ϕ ed/is ϕ -ing is that p

Dancy's thought is that if one thinks that the reasons for which the agent ϕ s are always facts about what the agent believes, then one will have to think of (1)-(3) as being elliptical for:

- (1a*) S ϕ ed/is ϕ -ing for the reason that they believe that p
- (1b*) The reason for which S ϕ ed/is ϕ -ing is that they believe that p
- (2a*) S ϕ ed/is ϕ -ing on the basis of the fact that they believe that p
- (2b*) The basis on which S ϕ ed/is ϕ -ing is that they believe that p
- (3a*) S ϕ ed/is ϕ -ing on the grounds that they believe that p
- (3b*) The grounds on which S ϕ ed/is ϕ -ing is that they believe that p

The problem is that (1*)-(3*) are counter-intuitive in a way that the following example brings out:

⁹There is another argument to be found in Dancy against the state version of Psychologism (Dancy, 2000:103-106). According to that argument, when the agent ϕ s for a good reason, the reason for which the agent ϕ s is identical to the normative reason in response to which they are ϕ -ing. It follows that the reasons which motivate the agent to ϕ must be the right sort of things to be normative reasons – they must occupy an ontological category such that, in so far as they belong to that category, it is possible for them to be normative reason. But normative reasons are *facts* whereas if the state version of Psychologism is true, the reasons which motivate would be *states*. Since states are not the right sort of thing to be fact, it follows that the reasons which motivate cannot be psychological states. I do not comment on this argument in the main body of the text. I do, however, think it's defensible and it's worth noting that the argument of the last chapter commits me to Dancy's claim that the reason for which one ϕ s in the good case is identical to the normative reason in response to which one ϕ s.

Consider a case where my reason for acting is genuinely that I believe that p. For instance, that I believe that the cliff is crumbling is my reason for avoiding climbing it, because having that belief I am more likely to fall off (I will get nervous). This is a case where that I believe what I do is genuinely my reason for action... But this is a quite unusual situation, not at all the normal case. Normally, I suppose that if things are not as I believe them to be, I do not in fact have the reason that I take myself to have. It would be quite peculiar to suppose that no practical reasons are like this, and that all are of the special sort that we found in the case of the crumbly cliff. (Dancy, 2000:124)

Dancy's plausible suggestion is that at the intuitive level, (1*)-(3*) are not true in ordinary contexts. This is brought out by the plausible thought that one only ϕ s for the reason that one believes that p in the special circumstances in which one treats the fact that one believes that p itself as a reason in favour of one's ϕ -ing. That we recognise these circumstances to be unusual is what brings it out that, intuitively, it is not the case that one's reason for ϕ -ing is always the fact that one believes that p. But the fact version of Psychologism is committed to saying that the (1*)-(3*) are true in ordinary contexts. So the fact version of Psychologism is committed to the denial of something that's intuitively compelling. Thus the onus of proof is on the proponent of fact Psychologism.

The argument against fact Psychologism just rehearsed can be readily generalised to state Psychologism, so that it is effective against Psychologism in general and delivers us a powerful reason to be Anti-Psychologistic. To achieve the generalisation, we'd have to replace the that-clause which appears in (1*)-(3*) with a phrase that picks out the corresponding psychological state of the subject: 'S's belief that p', or 'S's believing that p'. We could then run the same example to show that the resultant sentences are counter-intuitive in ordinary contexts, which would in turn shift the burden of proof onto state Psychologism.

Now, the argument so far has one significant dialectical feature: it is not decisive. It is an argument which serves to shift the burden of proof onto Psychologism. In §4.2 I will be attempting to undermine the best arguments in favour of Psychologism. So I could rest content with an onus shifting move here. But instead I want to develop the argument already rehearsed into a decisive argument in favour of Anti-Psychologism and leave the onus shifting argument of Dancy in the background. The resultant decisive argument is what I'm going to call the *Counter-Intuitiveness Argument*.

Dancy's plausible suggestion is that at the intuitive level, (1*)-(3*) are not true in ordinary contexts, as we have seen. This intuition is to be elicited by noting the specialness of the crumbly cliff case and its ilk. But what I want to suggest now is that the intuition that (1*)-(3*) are false in ordinary contexts can be explained only by appeal to the claim that the sort of relation involved in ϕ -ing for a reason is a relation which requires that the reason is an item which gets taken to be a normative reason by the subject. But it follows

from that claim that the reason for which one ϕ s is always to be understood as what one believes, for it is what one believes which is taken to be a normative reason by one. Only the Anti-Psychologist can say this, so Anti-Psychologism must be true.

This argument does not say: (1*)-(3*) are intuitively false in ordinary cases, but the proponent of Psychologism is committed to their truth, so the onus of proof is on the proponent of Psychologism. That's Dancy's own argument, as I understand it, generalised to all varieties of Psychologism. That argument is one which I think is ultimately effective against Psychologism, given my refutation of the main arguments for Psychologism to come in §4.2, but it is not to be confused with the Counter-Intuitiveness Argument. The Counter-Intuitiveness Argument is a deductive argument the conclusion of which is that Anti-Psychologism is true. It is not a burden-shifting argument which gets converted into a decisive argument when a refutation of the main counter-arguments is added. The Counter-Intuitiveness Argument says: (i) (1*)-(3*) are counter-intuitive in ordinary contexts; (ii) this can be explained only by appeal to the thought that the relation which holds between the subject and the reason for which they ϕ requires that the reason in question is taken to be a normative reason by the subject; hence: (iii) the reason for which the subject ϕ s is what gets taken by them to be a normative reason. Since it is only the Anti-Psychologist who can say this, it follows that Anti-Psychologism is true.

It's possible for the proponent of Psychologism to respond to the Counter-intuitiveness Argument by appeal to the Pryor semantics, or rather to the (1b)-(3b) analogue of it. The response here would be that Psychologism isn't committed to thinking that (1b)-(3b) are elliptical for (1b*)-(3b*). That's because (1b)-(3b) are not in the business of identifying the reason for which one ϕ s with the referent of any relevant that-clause. Rather, p functions to specify the content of the mental state that's referred to using the phrases such as 'the reason/basis/grounds for which/on which...' which appear in (1b)-(3b). Since (1b)-(3b) have the same meaning as (1a)-(3a), the latter are elliptical for their psychologistic counterparts only if the former are. Hence, (1a)-(3a) are not so elliptical either.¹⁰

The problem with this response is, of course, that the Pryor semantics is implausible and is, in any case, a theory which should not be our default view in this area. So the undermining of the Pryor semantics achieved previously has work to do both with respect to the Linguistic Argument and with respect to the Counter-Intuitiveness Argument.

This concludes my presentation and defence of the Counter-Intuitiveness Argument. I want now to end by drawing out two significant consequences of it.

First of all, the counter-intuitiveness of (1*)-(3*) in ordinary contexts and the explanation of why they are false in ordinary contexts offered here are independent of whether or not the factive reading of (1)-(3) is true. Let's suppose that the non-factive reading of (1)-(3) is true, so that they can be true even in the bad case. And let's suppose that a sentence of form (1) is indeed true of a subject in the bad case. The suggestion I make is

¹⁰This response would also serve to undermine Dancy's own onus shifting argument.

that even if this were so, it would be counter-intuitive to suggest that the reason for which the subject ϕ s is *really* that they believe that p, instead of just that p. And this intuition could still only be explained by appeal to the idea that the sort of relation between ϕ -ing and reasons must be one which requires the subject to take their reason to be a normative reason. The upshot is that even if one is impressed with the non-factive reading of (1)-(3), one will still have to accept that they concern a relation between the subject's ϕ -ing and an item which the subject takes to be a normative reason for them to ϕ . It's just that one would think of the relation in question as one which doesn't necessarily link the subject's ϕ -ing with a genuine normative reason.¹¹

Here is the second consequence I want to note. Earlier, I suggested that sentences of the form: 'S ϕ s for the reason that p' mean the same as sentences of the form: 'S ϕ s because p', in the rationalising sense of 'because'. In short, it is intuitive that (1)-(3) mean the same as non-psychologistic rationalising statements. It's a consequence of Dancy's claim, endorsed here, that (1*)-(3*) are counter-intuitive applied in ordinary contexts, that 'S ϕ s for the reason that they believe that p' does not mean the same as: 'S ϕ s because they believe that p', in the rationalising sense of 'because'. In other words, it is a consequence of Dancy's counter-intuitiveness claim that (1*)-(3*) do not mean the same as psychologistic rationalising statements. That's because (1*)-(3*) are intuitively false in all but the special circumstances in which the fact that one believes that p is itself taken to be a normative reason by one. But in ordinary circumstances psychologistic rationalising statements are true of one. Across ordinary and extra-ordinary cases, as well as across good cases and bad, it's true to say of a subject who ϕ s for a reason that they are ϕ -ing because they believe that p, in the rationalising sense of 'because', where p is whatever they take to be a reason in favour of their ϕ -ing. So if (1*)-(3*) are false in ordinary cases, yet in those cases there are corresponding psychologistic rationalising statements which are true, we get the result that the former cannot mean the same as the latter.

In the special circumstances in which it is true to say of someone that they ϕ for the reason that they believe that p, it will be true to say of them that they ϕ because they believe that p, in the rationalising sense of 'because'. But that rationalising statement is actually an instance of the non-psychologistic form of rationalising statement, for it takes as its apparent *explanans* what the agent takes to be a normative reason, instead of the state/fact of belief which constitutes the presence to mind of that which the agent takes to be a normative reason. So it remains true to say in general that sentences of the form 'S ϕ s for the reason that they believe that p' don't mean the same as their counterpart *psychologistic* rationalising statements.

Another way of putting the result here would be to say that (1)-(3) and (1*)-(3*) are asymmetric with respect to the relation each stands in to non-psychologistic and psycholo-

¹¹If we add to the mix here the claim that the relation in question is explanatory and non-causal what we have just is Dancy's View.

gistic rationalising statements respectively. The former mean the same as the corresponding class of rationalising statements. The latter don't. It follows from this that one of the elements of the picture argued for by Hyman in *How Knowledge Works* cannot be correct. Hyman focuses on a case discussed by Ryle (2000) in which a subject keeps to the edge because the ice is thin, in the rationalising sense of 'because'.¹² Ryle contrasts that success case with a case in which the subject keeps to the edge merely because they believe that the ice is thin. Hyman interprets Ryle as committing himself to the claim that in the success case, the subject keeps to the edge for the reason that the ice is thin whereas in the bad case they keep to the edge for the reason that they believe that the ice is thin: in the good case an instance of (1) is true of the subject whereas in the bad case an instance of (1*) is true of them.¹³ Hyman makes the following comment about Ryle's view, so interpreted:

Exactly so. In the first case, the man's reason is that the ice is thin; in the second case, it is that he believes that the ice is thin. But the man's reason stands to his action in a different relation in the two cases, and the sort of explanation given by identifying his reason differs commensurately. (Hyman, 1999:446)

An element of Hyman's position, at least in *How Knowledge Works*, then, is this. There are two forms of rationalising explanation: a good case form where the normative reason that *p* is *S*'s rationaliser and a neutral form where the fact that *S* believes that *p* is *S*'s rationaliser. We can report the first form by saying: '*S* ϕ s because *p*' and also by saying: '*S* ϕ s for the reason that *p*', where those sentences mean the same thing. We can report the second form by saying: '*S* ϕ s because they believe that *p*' and also by saying: '*S* ϕ s for the reason that they believe that *p*', where those two sentences mean the same thing too. Thus, for Hyman, ϕ -ing for the reason that *p* involves a different type of relation than ϕ -ing for the reason that one believes that *p*. The first is an explanatory relation that links a normative reason with a ϕ -ing where the reason in question is taken to be such by the subject. The second links a psychological fact to a ϕ -ing, where that psychological fact is designated as a reason, but is not taken to be a normative reason by the subject.

¹²In fact, the good case explanation in question which Ryle focuses on is an instance of: '*S* ϕ s because they know that *p*', but we can ignore that detail for our purposes in this chapter.

¹³What Ryle actually says about his case is this:

...to say that [someone] keeps to the edge because he knows that the ice is thin, is...to give quite a different sort of 'explanation', from that conveyed by saying that he keeps to the edge because he believes that the ice is thin. (Ryle, 2000:129-130)

Here Ryle doesn't commit himself to the claim that an instance of (1) is true in the good case and an instance of (1*) is true in the bad, only that an instance of '*S* ϕ s because they know that *p*' is true in the good case and an instance of '*S* ϕ s because they believe that *p*' is true in the bad. Hyman's interpretation of passage from Ryle reflects his taking it for granted that those explanatory sentences mean the same as their corresponding (1) and (1*) counterparts.

In Chapter Five I argue that Hyman is quite right to suggest that there are these two kinds of rationalising explanation, the first of which can be reported using a non-psychologistic rationalising statement, the second a psychologistic rationalising statement. But there are two thoughts I've just attributed to Hyman which cannot be right, given what's been said here. First, it cannot be right that 'S ϕ s for the reason that they believe that p' means the same as 'S ϕ s because they believe that p', for the latter is true in ordinary contexts but, as Dancy points out, the former isn't. Second, it cannot be right that 'S ϕ s for the reason that p' and 'S ϕ s for the reason that they believe that p' designate different types of relation. Dancy's point that (1*)-(3*) are counter-intuitive in ordinary contexts tells against the thought that the latter is true in ordinary contexts, and, as the Counter-Intuitiveness Argument has it, the only way to explain this is to say that the sort of relation specified by 'S ϕ s for the reason that they believe that p' is one which requires S to take whatever the reason is to be a normative reason. Hence, the relation involved in ϕ -ing for the reason that one believes that p cannot be a relation that's different from that which is specified by 'S ϕ s for the reason that p', for the first requires that the reason is what gets treated as a normative reason just as the second does.

In sum, what we should say about Ryle's case is that in the good case, he ϕ s because p and he ϕ s for the reason that p, where these are to say the same thing. In the bad case, he ϕ s because he believes that p, but he does not ϕ for a reason at all: there is no sentence 'S ϕ s for the reason that...' which is true of him, although there might be similar looking sentences which are, such as instances of (4)-(6) and (7), on their non-factive readings.

4.2 Two Arguments for Psychologism Rejected

I've defended two arguments for Anti-Psychologism. But there are several arguments in the literature in favour of Psychologism. If I can't say anything by way of undermining the latter, then my achievement in this chapter will merely be to plunge us into a state of *aporia* with respect to the Psychologism Question. This section examines the two most troubling arguments for Psychologism and rejects each of them. In §4.2.1 I consider an argument which appeals to the thought that the non-factive reading of (1)-(3) is correct and in §4.2.2 I consider an argument which appeals to the idea that we are prone to describe the reason for which we ϕ using the language of belief, in cases in which we doubt that p. That final argument also constitutes a fresh argument for the non-factive reading of (1)-(3).

4.2.1 The Argument from Error

The first argument for Psychologism I wish to reject I call the *Argument from Error*. Let's take a subject in the good case who ϕ s for the reason that p and let's focus on a correspond-

ing bad case in which p is false.¹⁴ The argument is this:

- (P4.1) The reason for which S ϕ s is the same in the good case and bad
 - (P4.2) In the bad case, p is false
 - (P4.3) The rationaliser of S's ϕ -ing when S ϕ s for a reason is identical to the reason in question
 - (P4.4) Explanations why can only take facts as *explanantia*
 - (P4.5) The reason for which S ϕ s is either: (i) the fact that S believes that p or (ii) p
 - (P4.6) The reason for which S ϕ s in the bad case must be a fact (From (P4.3) & (P4.4))
 - (P4.7) P cannot be the reason for which S ϕ s in the bad case (From (P4.2) & (P4.6))
 - (P4.8) The reason for which S ϕ s in the bad case is the fact that they believe that p (From (P4.5) & (P4.7))
-

- (C4.1) The reason for which S ϕ s in both good cases and bad is the fact that they believe that p (From (P4.1) & (P4.8))

In essence, the argument here is that the subject ϕ s for a reason in both good cases and bad, and it is the same reason in each case. This raises the question of what that reason is. It looks like there are two options: either the fact that S believes that p, or p itself. But since the reason for which one ϕ s plays an explanatory role, and reasons why must always be facts, p cannot be the reason for which S ϕ s in the bad case, for in the bad case p is false (or at least that's so with respect to the sorts of bad cases the argument focuses on). So in the bad case the reason for which one ϕ s must be that one believes that p. Our initial assumption was that the reason for which one ϕ s is the same in both good cases and bad. So the reason in question must be that one believes that p, and not p itself, in both cases.

One thing to note immediately about this argument is its conclusion. The conclusion is essentially that fact Psychologism is correct. The argument does not purport to prove state Psychologism correct, or to leave it open which variety of Psychologism is correct. That's still troubling for the proponent of Anti-Psychologism, of course.

(C4.1) is validly inferred from (P4.1) and (P4.8). Thus, the argument is valid. (P4.6)-(P4.8) are validly inferred from prior premises in the argument. So our question is: are each of (P4.1)-(P4.5) true? Our attention is restricted to bad cases in which p is false, so (P4.2) is trivially true. (P4.5) should also be granted, for Psychologism and Anti-Psychologism

¹⁴See Dancy (2000:121) for a statement of an argument for Psychologism which I take to be an underdeveloped version of the argument described here.

are the only options, and given that the argument presupposes that facts are *explanantia*, the fact version of Psychologism is the only Psychologistic option available. Finally, I'm granting that explanation why is factive here, so I grant (P4.4).

This just leaves (P4.1) and (P4.3). My view on (P4.3) is complicated, as was intimated in the introduction, and I haven't yet begun to develop it. So I'll leave (P4.3) aside. That just leaves (P4.1), and it is the obvious premise to reject. There are two ways of rejecting it. On the one hand, one might agree that $S \phi$ s for a reason in both the good case and bad, but say that the reason is different in each case: it's that p in the good case, but that one believes that p in the bad case. That way, one can agree that one's reasons are always facts, given the explanatory role they play, without being forced to accept a Psychologistic conclusion across the board. One would accept Psychologism about the bad case but Anti-Psychologism about the good case. The second way of rejecting (P4.1) would simply be to reject the claim that $S \phi$ s for a reason in the bad case at all.

It should be obvious which option I prefer: I effectively argued for the second option at length in the last chapter.¹⁵ The first premise of the argument goes wrong at a fundamental level: it makes the presupposition that $S \phi$ s for a reason at all in the bad case. That presupposition is mistaken, as we've seen. It should now be clear why the work of the last chapter is quite significant: it enables me to undermine the quite influential Argument from Error for Psychologism.

4.2.2 The Argument from Doubt

The final argument I want to consider for Psychologism is the *Argument from Doubt*. The argument in fact purports to establish two conclusions, both of which are troubling for me. The first conclusion is that the non-factive reading of (1)-(3) is true. The second conclusion is that Psychologism is true. Thus, the Argument from Doubt constitutes a problem both for my Anti-Psychologism as well as a fifth, and final, argument in favour of the non-factive reading of (1)-(3).

What does the argument say? Consider a case in which one ϕ s for the reason that p and one is asked why one is ϕ -ing. One will typically respond by saying: ' p ' or with a cognate sentence, so that one's answer to the question is that one is ϕ -ing for the reason that p . The Argument from Doubt says that in cases in which one has been made to question p by one's interlocutor one will move from answering the 'what's your reason for ϕ -ing?' question by asserting p to answering it by saying 'I believe that p '. Thus, the argument continues, the answer one gives in such cases is that one is ϕ -ing for the reason that one believes that p . Moreover, *it will seem to one as if one is offering precisely the same answer each time, just using different forms of words*. But since the second answer is clearly non-factive, the

¹⁵In any case, the first option cannot be right given that it is counter-intuitive that the agent ever ϕ s for the reason that they believe that p in ordinary contexts, even if one grants that they ϕ for a reason in bad cases. This was one of the lessons drawn from the Argument from Counter-Intuitiveness rehearsed above.

former must be too: the non-factive reading of (1)-(3) is correct, after all. Moreover, since the second answer implies that the reason for which one ϕ s is the state or fact of one's believing that p, we get the result that Psychologism is true.

For example, suppose Smith is asked for the reason for which he's walking to the quad. He would respond: 'there's free coffee available there'. But suppose further that Smith were prompted to doubt the claim that there's free coffee there in some way. At that point, he'd revert to saying: 'well, I *believe* there to be free coffee available there'. The thought is that the first response commits Smith to the claim that the reason for which he walks is that there's free coffee available whilst the second commits him to the claim that he walks for the reason that he believes there to be free coffee there. Moreover, it seems to Smith that he's offering the same answer each time. But since the second answer doesn't require it to actually be the case that there's coffee in the quad and identifies Smith's reason with his belief, we have to conclude that, really, the claim that Smith walks to the quad for the reason that there's free coffee available there is non-factive and is elliptical for a corresponding instance of (1a*).

Let's look at the argument in more detail. What the argument focuses on is a certain type of conversational exchange which is directed towards the question of what the reason for which one is ϕ -ing is. One would, ordinarily, respond to the question by asserting the truth of p – which is what one takes to be a normative reason in favour of one's ϕ -ing. The conversational exchange in question involves one's interlocutor prompting one to question whether p really is true – they prompt one to doubt whether p really is the case. The prompt might involve the presentation of reasons for thinking that p is false, or it could take some other form, for example the simple assertion that p isn't actually the case. Whatever form the prompt to question p takes, in response to the prompt one would respond by casting one's answer in psychologistic language – one would move from asserting p, to saying something along the lines of: 'well, alright then: I *believe* that p'.

There are then two claims made about the exchange by the proponent of the argument. The first claim is that pre-doubt one is committing oneself to the following answer to the question: *the reason for which I ϕ is that p* and post-doubt one is committing oneself to the following, psychologistically expressed, answer to the question: *the reason for which I ϕ is that I believe that p*. The second claim about the sort of exchange made by the proponent of the argument is that, from one's own point of view, the two forms of words one uses pre and post-doubt express the very same answer to the question. As Dancy (2000), following Collins (1997), would put it: the psychologistic answer to the question – 'I believe that p' – seems to one to be a *re-statement* of the non-psychologistic statement – 'p'. It doesn't seem to one as if one is withdrawing the thought expressed by the first statement and then offering one's interlocutor a quite different thought, expressed by the second. Rather: the thought offered is the same, but is offered using different forms of words. Putting those two claims together, we get the result that from one's own point of view the answer to the

question which dovetails with the simple assertion of p , namely: that the reason for which one is ϕ -ing is that p , is the very same answer to the question as the answer which dovetails with the psychologistic response ‘I believe that p ’, namely: that the reason for which one is ϕ -ing is that one believes that p .

Given that we shouldn’t convict ordinary subjects with systematic error with respect to what their answers are to such questions, we must take it that the two answers really are the same. There are then two further inferences made from this: an inference to the non-factivity of (1)-(3) and an inference to Psychologism. First, the inference to the non-factivity of (1)-(3). The thought there is that since the second, psychologistic, answer is non-factive, so must be the first, non-psychologistic answer. The resulting view is the sort of conception of ϕ -ing for reasons which is associated with the Motivating Reasons View and which should already be familiar. Second, the inference to Psychologism. The thought is that since the psychologistic answer identifies the state or fact of the subject’s belief that p as the reason for which they ϕ , and the psychologistic answer is the same as the non-psychologistic answer, the reason for which the subject ϕ s must be their state/fact of believing that p . ϕ -ing for the reason that p is always really ϕ -ing for the reason that one believes that p . It takes reflection on cases of doubt to bring this out. (1)-(3) really are elliptical for (1*)-(3*).

In order to make a start on undermining the Argument from Doubt I first want to suggest that there is an important distinction to be drawn between two variants of the basic sort of conversational exchange in question. First, there is the variant in which the subject continues to believe that p , despite being prompted to doubt that p by their interlocutor. In that variant of the basic sort of case, the subject holds fast to their belief despite the prompt to doubt offered by their interlocutor, perhaps putting off engaging in self-critical reflection about it for a later date, or perhaps agreeing with whatever objections have been raised but only lowering credence slightly in response to them. Second, there is a variant of the case in which the subject drops their belief that p in response to the prompt to doubt. In that variant of the basic sort of case, the subject agrees with their interlocutor and ends up exchanging their belief that p for a suspension of judgement about p or else an outright belief that not- p . In that variant of the basic sort of case, what will be at issue is not a transition from ‘ p ’ to ‘I believe that p ’, but a transition from ‘ p ’ to the past-tense ‘I believed that p ’.

With that distinction in the background the proponent of the argument will need to show that with respect to *at least one variant* of the basic sort of case, the Argument from Doubt holds. What I want to do now is suggest that the argument fails for each variant. Regarding the first variant, my main contention will be that the argument doesn’t go through because there is a special use of ‘I believe that p ’ operated with by the subject in such cases on which the subject is simply using the phrase in order to hedge their assertion that p . With respect to the second variant, my contention will be that the answer to the

question one is giving post-doubt is not really 'I ϕ ed for the reason that I believed that p', but 'I ϕ ed because I believed that p', which, as we have seen, is different.

I'll start with the first variant, in which one continues to believe that p post-doubt. With respect to the first variant, the initial thought would have to be that prior to doubt, one asserts: 'p', post-doubt one says: 'I believe that p', *but one continues to believe that p post-doubt*. The answer one gives pre-doubt is that the reason for which one ϕ s is that p; the answer post-doubt, where one is continuing to believe that p, is that the reason for which one ϕ s is that one believes that p. The further thought would then be that in such a situation, it seems to the subject as if the psychologistic idiom expresses the same answer to the question of what the reason is as the non-psychologistic idiom. And the final thought would be that the psychologistic answer doesn't require the truth of p and identifies the agent's state/fact of believing as the reason for which they ϕ .

In response, I want to first of all examine some remarks made about belief-talk by Bernard Williams in *Deciding to Believe*:

...to say 'I believe that p' itself carries, in general, a claim that p is true. To say 'I believe that p' conveys the message that p is the case. It is a way, though perhaps a somewhat qualified way, of asserting that p is true. (Williams, 1970:137)

Elaborating on this somewhat, Williams goes on to say this:

The most elementary and straightforward expression of the belief that it is raining is to say 'it is raining', not to say 'I believe that it's raining'. 'I believe that it's raining' does a rather special job. As a matter of fact, it does a variety of special jobs. In some cases, it makes what is very like an autobiographical remark; but very often in our discourse it does a special job of expressing the belief that p, or asserting the belief that p, in a rather qualified way. On the whole, if somebody says to me, 'Where is the railroad station?' and I say 'I believe that it's three blocks down there and to the right', he will have slightly less confidence in my utterances than if I just say 'It's three blocks down there and to the right.' (*Ibid.*:137-138)

What's happening here is that Williams is seeking to draw our attention to two uses of the phrase 'I believe that p'. On the one hand, there's a 'reporting' use of the phrase, in which one is attributing to oneself a state of a certain type, where what one says is true just in case one is in a state of that type. On the other hand, there's an 'expressive' use of the phrase, in which one is evincing one's state of believing that p in the same sense in which one does so when one simply asserts that p. The thought is that on the expressive use, 'I believe that p' just is an assertion that p, but where one is hedging one's endorsement of p, or expressing a certain sort of lack of confidence in what one is asserting.

So even though asserting that p is the paradigm way of evincing one's belief that p , a phrase which is sometimes used to report the fact that one is in a state of believing can also be used to evince one's belief. A good analogy here is with the phrase 'I want x '. That phrase can be used to report the fact that one is in a certain sort of pro-attitude state, or it can be used to express that very state which one is in.¹⁶

Relying on this material from Williams, it is possible to undermine the Argument from Doubt, applied to the first variant of the basic type of conversational exchange we've been focusing on. For one continues to believe that p even post-doubt in the first variant of the case. This makes it plausible that the phrase 'I believe that p ' as one uses it post-doubt, is functioning not to report that one is in a state of believing that p but to express one's belief that p , albeit in the hedged way Williams describes. With that point in tow, it can now be said that the answer to the question one commits oneself to post-doubt is indeed the same as the answer to the question pre-doubt, but the answer to the question is just the non-psychologistic one throughout: *I'm ϕ -ing for the reason that p* . It's just that, post-doubt, one is hedging the assertion of p one makes in so far as one asserts that sentence. Thus, no conclusion in favour of the non-factive reading of (1)-(3) or Psychologism follows.

Now I'll move on to examining whether the argument might work when applied to the second variant of the basic sort of case, in which one drops one's belief that p post-doubt. With respect to that second variant, the initial thought would have to be that prior to doubt, one asserts: ' p ', post-doubt *one has dropped one's belief that p* , so one asserts the past-tense: 'I believed that p '. The first response is supposed to commit one to the claim that one is ϕ -ing for the reason that p ; the second to the claim that one was ϕ -ing for the reason that one believed that p . The further thought will be that from one's own point of view, the second response is merely a past-tensed way of reporting the very same answer to the question as the first: each response characterises one's reason for ϕ -ing in the same way, the only difference is the tense in which things are reported. The final thought will then be that since 'I ϕ ed for the reason that I believed that p ' doesn't require that p was true at the time of one's ϕ -ing and because it identifies the reason at the time of one's ϕ -ing with a state/fact of belief, so the 'I'm ϕ -ing for the reason that p ' answer must likewise be non-factive and identify the reason for which one ϕ s with a state/fact of belief.

In response to this, it's worth offering the reminder that it is quite counter-intuitive in ordinary contexts to identify one's reason with the state/fact of one's belief that p . Only in Dancy's special contexts is it intuitive to identify one's reason with such a factor. That goes for past as well as present tense characterisations of one's reasons. In the sorts of cases at issue, I suggest, when one responds post-doubt by saying 'I believed that p ', the answer to the question one is committing oneself to is not that one's reason for ϕ -ing was that one believed that p , but that one was ϕ -ing because one believed that p , in the rationalising sense of 'because'. But if that's what's going on, then it's false that from one's own point

¹⁶Compare Marcus (2012:§1.3).

of view the answer provided by the pre-doubt response is the same as the answer provided by the post-doubt response. The former says that one is ϕ -ing for the reason that p. The latter retracts that claim in favour of the claim that one was ϕ -ing because one believed that p. Only in the first variant of the basic sort of conversational exchange, in which one continues to believe that p post-doubt, is there pressure to say that it seems to one as if one's answer is the same each time.

This concludes my response to the Argument from Doubt. In sum, we've seen that there are contexts in which one is prompted to question the truth of what one takes to be a normative reason for one's ϕ -ing by an interlocutor, and in those contexts one is disposed to make a transition from characterising one's reason in a non-psychologistic manner to characterising it in a psychologistic manner. If one retains the belief post-doubt, one isn't saying that one is ϕ -ing for the reason that one believes that p, when one makes the transition. Rather, one is saying that one is ϕ -ing for the reason that p, but it's just that one is using belief-talk to hedge one's assertion that p, in the familiar way Williams describes. If, on the other hand, one drops one's belief, then one will resort to a past-tense characterisation of one's ϕ -ing, and this would not seem to one to be simply a past tensed way of recording the very same thing that one records when one says 'I'm ϕ -ing for the reason that p'. Rather, it would be matter of explaining why one ϕ ed by appeal to one's belief that p.

It is worth noting, by way of finishing, that the Argument from Doubt would not work were it to concern linguistic contexts in which one is describing the reason for which *someone else* is ϕ -ing, instead of first-personal contexts in which the agent who ϕ s for the reason that p herself is subject to the doubt and so re-states things psychologically. In the third-personal form of the case, the proponent of the argument would have to say that one would move from describing someone else as ϕ -ing for the reason that p to describing them as ϕ -ing for the reason that they believe that p, in response to the sort of doubt about p put to one by one's interlocutor. The problem with this third-personal form of the Argument from Doubt is the same as the problem for the version of it which focuses on first-personal cases in which one drops one's belief post-doubt. It is quite counter-intuitive to characterise an agent's reason for ϕ -ing as a state/fact of believing, in ordinary contexts. That goes whether or not the agent in question is oneself or someone else. In third personal contexts what one would do is characterise the agent in question as ϕ -ing because they believe that p, which, as we've seen, is quite different from saying that their reason for ϕ -ing is that they believe that p.

4.3 The Explanatory Question

In the last chapter I reconfigured the Dialectic of Psychologism in light of the fact that there are no such things as motivating reasons in the special sense described. The Dialectic of

Psychologism is not to be understood as a debate about the nature of motivating reasons thought of as entities that count as reasons and which play a rationalising role across both the good case and bad: whether they are psychological items and whether they are identical to rationalisers.

I take myself to have articulated and defended an answer to the Psychologism Question, which is one half of the Dialectic of Psychologism. My answer is this: the reasons for which the agent ϕ s are what the agent takes to be the normative reasons in favour of their ϕ -ing.¹⁷ They are not the beliefs the agent has which constitute the presence to mind of such normative reasons, nor facts which record the existence of such states. The proponent of Psychologism falls foul of intuitions concerning (1*)-(3*) as well as powerful linguistic considerations which tell in favour of Anti-Psychologism. The arguments for Psychologism described are, moreover, all ineffective.

Along the way, a number of interesting results have emerged. First, that ‘S ϕ s for the reason that S believes that p’ is false in ordinary contexts and so cannot be equivalent to ‘S ϕ s because S believes that p’ in those contexts: the former is not simply another way of stating the latter. Second, that there are contexts in which one uses psychologistic language to characterise one’s reason, but those contexts involve hedged assertions that p or else offer psychologistic rationalising explanations. Third, the refutation of the Motivating Reasons Postulate has helped to undermine Psychologism by supplying a basis for a response to a would-be powerful argument for Psychologism.

All of that is by way of summary of my response to the first part of the Dialectic of Psychologism. This section makes a start on the second part of the Dialectic of Psychologism: the Explanatory Question, which asks: when the agent ϕ s for a reason, is the *explanans* of the rationalising explanation to which their ϕ -ing is subject to be identified with the reason for which they ϕ ? As we shall see, that second question cannot be given a definitive answer until the next chapter, and so the Dialectic of Psychologism cannot be fully resolved until then either. The burden of this final section is to set up my answer to the Explanatory Question by sketching the solution I want to provide and fitting it into the dialectic as it has been presented so far.

The Dancy of *Practical Reality* and Davidson are committed to a positive answer to the Explanatory Question. The Neo-Davidsonians are committed to a negative answer. How should we answer the question?

I will be arguing for a positive answer to the question: whenever one ϕ s for a reason, the reason for which one ϕ s is one’s rationaliser. My argument for a positive answer hinges on the argument, to be given in the next chapter, that there is an essentially successful form of rationalising explanation which takes the fact that p *qua* normative reason as *explanans* and which obtains if, and only if, one is responding to a normative reason, as well as

¹⁷Indeed, given the results of the last chapter: the reasons which motivate are always genuine normative reasons, typically about the external world.

a neutral type of explanation which takes the fact that one believes that p as *explanans*. Since it was shown in the last chapter that when one ϕ s for the reason that p , the reason for which one ϕ s just is the normative reason to which one is responding, it follows that when one ϕ s for the reason that p , there is a rationalising explanation to which one's ϕ -ing is subject which takes the reason for which one ϕ s as *explanans*. That just is a positive answer to the Explanatory Question.

In light of my commitment to the thought that there are fundamentally two kinds of rationalising explanation, and also in light of my commitment to the claim that one ϕ s for a reason only in the good case, my positive answer to the question takes on a certain distinctive shape. I argue for a positive answer to the Explanatory Question. But given that I think that one only ϕ s for a reason in the good case I do not commit myself to the claim that all rationalising explanations take the reason for which the agent ϕ s as *explanans*: only the rationalising explanations particular to the good case do so. And because I think that there are two fundamental types of rationalising explanation where both hold in the good case (although one of them holds in the bad case too), I do not think that every rationalising explanation which holds in the good case has the agent's reason as its *explanans*: only the essentially successful type does. So my positive answer to the Explanatory Question is not marketed as a claim about the nature of rationalising explanation in general. The claim is that of all the cases in which one's ϕ -ing is subject to a rationalising explanation of any form, there is a sub-set of those cases in which one ϕ s for a reason. In that sub-set of cases one's ϕ -ing is subject to two different kinds of rationalising explanation: a neutral one and an essentially successful one. The view I defend is that the essentially successful one has an *explanans* identical to the reason for which one ϕ s.

It is worth bringing out how this kind of positive answer to the Explanatory Question differs from the positive answer provided by Dancy and Davidson. Dancy and Davidson agree that whenever one ϕ s for a reason, the reason for which one ϕ s is identical to one's rationaliser, it's just that Davidson suggests, but Dancy denies, that the reason in question is a psychological item. But they go further. They say that rationalising explanation always takes the reason for which one ϕ s as its *explanans*. That's because, given that they are proponents of the Motivating Reasons Postulate, they think that one counts as ϕ -ing for a reason across both good cases and bad. Good cases and bad exhaust the set of cases in which one's ϕ -ing is subject to a rationalising explanation. So when they agree that when one ϕ s for a reason, the reason for which one ϕ s is one's rationaliser, what they commit themselves to is the thought that whenever one's ϕ -ing is subject to a rationalising explanation, the explanation to which one's ϕ -ing is subject has the reason for which one ϕ s as its *explanans*. Moreover, as proponents of Reductionism about the good case, they think that there is only type of rationalising explanation. So, on their view, there cannot be a form of rationalising explanation present in either case which does take as *explanans* something different to the reason for which the agent ϕ s. In short: their positive answer to

the Explanatory Question, given their background commitment to the Motivating Reasons View, commits them to saying that something is a rationaliser if, and only if, it is a reason which motivates the agent to ϕ . That is not my position. I allow that there are rationalisers which are not reasons for which the subject ϕ s. The neutral explanation fits that bill. It's just that there are also rationalisers which are identical to the reasons for which one ϕ s: the essentially successful explanations fit that second bill.

It should be clear now how my positive answer to the Explanatory Question relates to the positive answer provided by Davidson and Dancy. But how does my positive answer fit in with the negative answer provided by the Neo-Davidsonians? The Neo-Davidsonians think that the only type of rationalising explanation there is is a neutral type which takes the state or fact of belief as *explanans*. This is explicitly supposed to be consistent with the thought that the reasons for which the subject ϕ s (which are conceived of as motivating reasons in the special sense by the Neo-Davidsonian) are non-psychological. So Neo-Davidsonianism, like Dancy's view but unlike the view of Davidson, is consistent with the main conclusion of this chapter. One obvious way in which my view differs from that of the Neo-Davidsonians is that I think that there is a non-neutral type of rationalising explanation. But I also think that that non-neutral type takes the fact that p, as opposed to some psychological factor, as *explanans*. Moreover, central to Neo-Davidsonianism is the thought that (1)-(3), in so far as they express a rationalising explanation, do not express an explanation with p as its *explanans* but with the agent's state or fact of believing that p as *explanans*. My view predicts, like Dancy's, that (1)-(3) report a rationalising explanation which does take p as its *explanans*. Despite all this disagreement, however, I agree with the Neo-Davidsonians that there is a type of rationalising explanation which is neutral, which takes the fact that one believes that p as *explanans*, and which doesn't have as *explanans* a factor that counts as the reason for which the agent ϕ s. My position thus represents something of a half-way house between the position of Davidson and Dancy on the one hand, and the Neo-Davidsonians on the other.

What motivates Neo-Davidsonianism is the thought that we need to avoid the claim that rationalising explanations are non-factive. Dancy explicitly runs into that problem because he takes it that one can ϕ for the reason that p even when not-p, that the reason for which one ϕ s when one ϕ s for the reason that p is p itself, that when one ϕ s for the reason that p there is a rationalising explanation which applies to one's ϕ -ing, and that p is the *explanans* of rationalising explanation. Thus, one's ϕ -ing can be subject to an explanation why which takes p as *explanans* even when not-p. Wishing to avoid this unpalatable consequence but also wishing to subscribe to the claim that ϕ -ing for the reason that p can happen even when not-p, that ϕ -ing for the reason that p involves one's ϕ -ing being subject to a rationalising explanation, and Dancy's Anti-Psychologism, the Neo-Davidsonians are motivated to deny that rationalisers are one's reasons.

The possibility of my own position on this matter undercuts the motivation for Neo-

Davidsonianism. More precisely, it points to a way to avoid the problem without opting for a negative answer to the Explanatory Question. The claim to reject is that one ϕ s for the reason that p even when not-p. That way, one can allow that ϕ -ing for the reason that p involves one's ϕ -ing being subject to a rationalising explanation, Dancy's Anti-Psychologism, that p is the *explanans* of rationalising explanation when one ϕ s for the reason that p, *and* factivity. Factivity is preserved because it is denied that one can ϕ for the reason that p even when p is false. Centrally, the position in question requires that the rationalising explanation present when one ϕ s for the reason that p is distinct in kind from that present in the bad case – a possibility which Neo-Davidsonians do not acknowledge but which holds the key to my solution to the Explanatory Question.

That is how I wish to answer the Explanatory Question. The argument for the theory of rationalising explanation which provides me with my answer to it will come in the next and final chapter of this essay, in which I explore the question of whether we should go Reductionist about the good case. So settling the Explanatory Question will have to be put off until that argument has been defended.

Chapter 5

Williams' Dictum

I've been operating with a distinction, introduced in Chapter One, between good cases and bad cases of taking p to be a reason for one to ϕ and ϕ -ing accordingly. A good case is a case in which the subject responds to the genuine normative reason that p . A bad case is a case in which the agent takes p to be a reason for them to ϕ and ϕ s in a way that manifests their taking p to be a reason in favour of their ϕ -ing, but they nevertheless blamelessly fail in responding to a normative reason. Most straightforwardly, this will be because p isn't really a reason for them to ϕ at all, even though they blamelessly take it to be, but it could be that p is a reason for them to ϕ , and yet they blamelessly fail to be connected to p in the right sort of way.

So far, I've been paying almost all of my attention to the condition of ϕ -ing for the reason that p . I've argued, in Chapter Three, that it just is the success condition of responding to the normative reason that p , so that it is a condition which does not obtain in the bad case. This refutes the popular Motivating Reason View, articulated in Chapter Two, according to which the subject ϕ s for reasons even in the bad case. I've also argued, in Chapter Four, against a Psychologistic construal of the reasons for which we ϕ .

All of this material can be, and at points has been, related to the notion of rationalising explanation, also introduced in Chapter One. A rationalising explanation is a type of explanation why which explains why the agent ϕ ed or is ϕ -ing by appeal, in some way, to what the agent takes to be a reason in favour of their ϕ -ing. Across good cases and bad cases, the agent's ϕ -ing is subject to a rationalising explanation which can be reported using a sentence of the form 'S ϕ ed/is ϕ -ing because S believes that p '. In the good case only, a rationalising explanation can also be reported using: 'S ϕ ed/is ϕ -ing because p '. The condition of ϕ -ing for the reason that p is connected to the notion of rationalising explanation: ϕ -ing for the reason that p is ϕ -ing in a way that is subject to a rationalising explanation which takes as its *explanans*, wholly or partly, the reason for which one is ϕ -ing.

Although I have brought into play the notion of a rationalising explanation on a number of occasions I have not said much about what our account of it should be, other than to

provide a brief sketch of the account I want to offer, and to bring out its relation to the Explanatory Question, at the end of the last chapter. This chapter aims to fill in that lacuna via the settling of an outstanding issue: whether or not we should go for a Reductionist account of what it is to respond to a normative reason.

My focus is going to be on the rights and wrongs of the following thesis, articulated once again by Williams:

... the distinction between false and true beliefs on the agent's part cannot alter the *form* of the explanation which will be appropriate to his action. (Williams, 1980:120)

I shall refer to the thesis committed to by Williams here *Williams' Dictum*. I will be suggesting that Williams' Dictum is ambiguous, but that on one reading it says that there is a single type of rationalising explanation and that type is neutral with respect to whether or not p is a genuine reason for S to ϕ . Contrary to Williams' Dictum, on that reading, I will be arguing that there are two different types of rationalising explanation: one which is neutral in the way described and which takes the fact that S believes that p as *explanans* on the one hand, and on the other hand one that has the fact that p *qua* normative reason as *explanans*, which can hold only in the good case. Refuting Williams' Dictum in the way described delivers us the result that Reductionism is false as well as delivering us the desired answer to the Explanatory Question, as we will see.

I will proceed as follows. In §5.1 I present a preliminary account of ϕ -ing in response to the normative reason that p , laying out some salient necessary conditions. In §5.2 I offer an interpretation of Williams' Dictum. I suggest that it is ambiguous, in the light of the distinction between a form of explanation and particular explanations which exemplify the form. In §5.3 I bring Reductionism onto the scene and relate it to Williams' Dictum on both of its readings. In §5.4 I present my own positive view, which constitutes a rejection of Williams' Dictum and hence Reductionism. In §5.5 I examine and reject two arguments currently available in the literature against Williams' Dictum. This creates a need for a fresh argument and in §5.6 I attempt to provide one by first demonstrating that my preferred view should be the default view in this area, thus shifting the burden of proof onto the proponent of Williams' Dictum and, second, rejecting extant arguments for Williams' Dictum. Finally, in §5.7 I address the question of whether we should think of the type of explanation which I argue is special to the good case as in some sense primary. I argue that my own position delivers the relevant primacy thesis.

5.1 The Elements of Success

I begin with an attempt to get clear about what is required for the subject to ϕ in response to a normative reason, which builds on the account offered in Chapter One. I am not

going to attempt to offer a set of individually necessary and jointly sufficient conditions for responding to a normative reason. Nor am I going to attempt to offer anything worth describing as a complete account of that condition. My intention is rather to get on the table some salient necessary conditions of responding to a genuine reason. Much of my energy will be spent defending the claim put on the map by Hornsby (2007a, 2007b, 2008) and Hyman (1999, 2006, 2010, 2011), and later defended by McDowell (2013b) and Marcus (2012), that responding to a normative reason requires knowledge.

Suppose Smith walks to the quad for the reason that there is free coffee available there. And suppose Smith's case is a good case: he is responding to a genuine normative reason. What is going on with Smith, in so far as he is responding to a normative reason?

The first thing to say is that Smith's action is subject to a rationalising explanation: the holding of a certain rationalising explanation either is, or is at least an element of, what it is to respond to a normative reason. An explanation of the relevant sort can be reported in myriad ways, as we have seen. The holding of the rationalising explanation will require, as we saw in Chapter One, certain psychological elements, for example that the subject believes that p and takes p to be a normative reason. It may also require more (for example: a causal relation between Smith's action and his belief), depending on one's theory of rationalising explanation.

A second thing we can say immediately is that responding to a normative reason requires the reason to which one is responding to actually be a reason for one to ϕ . It needs to be true that there is free coffee available in the quad and it needs to be the case that that has the normative force Smith takes it to have.

The two necessary conditions just broached are connected. The first condition says that a rationalising explanation applies to Smith's action. The second says that Smith's reason must be a genuine reason. If the normative reason is that p , then p , or some element corresponding to it, will have to at least be a part of the *explanans* of the rationalising explanation in question. The explanation to which Smith's ϕ -ing is subject will have to involve Smith's reason: that there is free coffee available in the quad, which, by the second condition, is a genuine one.

But ϕ -ing in a way that manifests one's belief that p and one's taking p to be a normative reason, so that one's ϕ -ing is subject to a rationalising explanation, does not suffice for responding to a normative reason, even when p really is a reason for one to ϕ . Something more is required, as Hornsby and Hyman point out, in order for one's ϕ -ing to be linked up to the normative reason in the way distinctive of the good case. To see what, let us focus on the following example, presented by Hornsby:

The example concerns Edmund, who believes that the ice in the middle of the pond is dangerously thin, having been told so by a normally reliable friend, and who accordingly keeps to the edge. But Edmund's friend didn't want Edmund to skate in the middle of the pond (never mind why), so that he told

Edmund that the ice there was thin despite having no view about whether or not it actually was thin. Edmund, then, did not keep to the edge *because* the ice in the middle was thin. Suppose now that, as it happened, the ice in the middle of the pond was thin. This makes no difference. Edmund still didn't keep to the edge *because* the ice was thin. (Hornsby, 2008:251)

What we have here is a case in which the subject is performing an action for which he believes there to be a decisive reason. The action is one of skating on the outskirts of the frozen pond and the consideration taken to be a decisive reason is: the ice in the middle of the pond is thin. Edmund justifiably believes, in some sense of 'justification', that the ice in the middle is thin and his source of justification is testimony from his normally reliable friend. Now we have two variants of that basic case: a case in which it's true that the ice is thin and a case in which it isn't. In the latter case, Edmund is mistaken in treating the relevant consideration as a decisive reason for action because normative reasons are facts. Moreover, it's clear that Edmund doesn't stick to the edge because the ice is thin: acting because *p* requires the truth of *p*. But more interesting is the former variant of the case, in which it is true that the ice in the middle is thin. Edmund has a justified true belief that the ice in the middle is thin, but he doesn't know it. The case is a type of Gettier Case. Moreover, it's intuitive that even in that situation, Edmund still doesn't act because the ice in the middle is thin. The plausible suggestion is that we can best explain why not by appeal to the thought that acting because *p*, in the rationalising sense of 'because', requires nothing short of knowing that *p*.¹

Hornsby's plausible conclusion is marketed as a claim about acting *because p*, in the rationalising sense of 'because'.² Since the condition reported by '*S* ϕ s because *p*' just is the success condition of responding to the normative reason that *p*, we can conclude that acting in response to the normative reason that *p* requires knowing that *p*.

Hornsby defends a claim which links knowing with *acting* in response to a normative reason. I think her argument is sound. Moreover, I think it can be generalised to cover all judgement sensitive phenomena, not just actions. That is, I think the following thesis is plausible:

The Epistemic Thesis. *S* ϕ s in response to the normative reason that *p* only if *S* knows that *p*.

¹McDowell (2013a) relies on the same case as Hornsby.

²This is why Dancy (2008) fails to undermine Hornsby's argument by suggesting that the intuition is not generated when the condition is not *S* ϕ -ing because *p*, but one of (1)-(7). As McDowell (2013b) points out, this objection fails because even if it's granted that the intuition is not generated with respect to the condition reported by (1)-(7), it doesn't follow that it's not generated with respect to the condition reported by '*S* ϕ s because *p*', which is the condition Hornsby focuses on. For the record, I think the intuition does hold for (1)-(3). It does not hold for (4)-(6) and (7) on the non-factive readings thereof, but it does hold for them on their factive readings.

Responding to the normative reason that p , so that one counts as ϕ -ing because p , requires knowing that p , whether the type of ϕ -ing in question is acting, omitting, believing, intending, hoping, doubting, imagining, desiring, or whatever.³

We've seen what can be said in favour of the Epistemic Thesis applied specifically to action. How might we lend support to the much stronger, generalised thesis? Hornsby's case can readily be extended to omissions as well as to what we might think of as the practical attitudes: intending, desiring, wishing, hoping, and so fourth. Moreover, I think we can appeal to a case offered by Hyman (1999) in order to prove that the thesis holds for belief. I assume that once it's been shown that the thesis holds for the central practical phenomena and that it holds also for belief, we can reasonably infer that the thesis can be extended to judgement sensitive phenomena across the board. Here is Hyman's case:⁴

Henry is watching the television on a June afternoon. It is Wimbledon men's finals day, and the television shows McEnroe beating Connors; the score is two sets to love and match point to McEnroe in the third. McEnroe wins the point. Henry believes justifiably that:

(1) I have just seen McEnroe win this year's Wimbledon final

and reasonably infers that

(2) McEnroe is this year's Wimbledon champion.

Actually, however, the cameras at Wimbledon have ceased to function, and the television is showing a recording of last year's match. But while it does so McEnroe is in the process of repeating last year's slaughter. So Henry's belief (2) is true, and surely he is justified in believing (2). But we would hardly allow that Henry knows (2). (*Ibid.*:447)

Now let's suppose, for example, that Henry believes that McEnroe is pleased with himself. Does Henry harbour that belief because McEnroe is this year's Wimbledon champion, in the rationalising sense of 'because'? I don't think so. The best explanation of this is that believing in response to a normative reason requires knowing it.⁵ Putting that result together with the previous results in the practical domain I think we get a strong case in favour of the Epistemic Thesis.

³Hyman supports the generalised thesis. Marcus supports the thesis applied to action *and* belief, and I see no reason why he'd wish to deny the generalised thesis too. Hornsby and McDowell restrict themselves to the action case only, although, again, I see no reason why they wouldn't also wish to go for the generalised thesis to be defended here.

⁴The case in question, as Hyman notes, was given its first written presentation by Dancy (1985), who in turn attributes it to Brian Garrett.

⁵See also Marcus (2012:37-38) for additional Gettier style cases which purport to prove the doxastic instance of the Epistemic Thesis.

I've offered an argument for Epistemic Thesis. But now I want to consider an objection to it raised by Hughes (2014). Hughes distinguishes between two sorts of Gettier cases: those in which there is a causal connection between the agent's belief that p and the fact that p and those which do not involve such a causal connection. The well-known *Barn Façade* Case is an instance of the former, whereas Hornsby's Edmund Case, Hyman's Tennis Case and Gettier's own cases are an instance of the latter. Hughes suggests that in Causal Gettier Cases, or at least in the Causal Gettier Cases he considers, the subject fails to know that p , but they are in a position to ϕ in response to the normative reason that p , nevertheless. This contradicts the Epistemic Thesis. Hughes considers two cases: a variant of Hyman's Tennis Case and a variant of the standard Barn Façade case. The latter is the simpler of the two, so I'll consider that one, although what I have to say in response to it applies equally well to the former:

Henry is out hiking. He's lost, and the weather is turning nasty. The situation is getting serious. He sees what he believes to be a hiker's hut in the distance, and feels relieved. In fact, unbeknownst to Henry, he is in fake hiker's-hut county – an area where there are only a handful of real huts, and many hut façades designed to look exactly like real huts to passing hikers. Henry justifiably and truly believes that the structure in the distance is a hut, but he does not know this. (*Ibid.*:461)

And here is what Hughes says about it:

Is the fact that there is a hut in the distance Henry's reason for being relieved? Intuitively, I submit, it is. (*Ibid.*)

The case is supposed to be one in which Henry truly believes that there is a hiker's hut in the distance, and that that is a reason for him to feel relieved. He feels relieved in a way that manifests those commitments. Henry doesn't *know* that there is a hiker's hut in the distance, however, by dint of his being in Hut-Façade County. That's so even though the fact that there is a hiker's hut in the distance causes Henry's belief. It is supposed to be the case, nevertheless, that Henry feels relieved in response to the normative reason that there is a hiker's hut in the distance. The supposed upshot is that the Epistemic Thesis is false: there are cases in which S ϕ s in response to the normative reason that p , but they do not know that p .

Now, there's an obvious response to this on the part of the proponent of the Epistemic Thesis which is, in the end, defensible. The response consists in the proponent of the Epistemic Thesis attempting to explain away Hughes' intuition consistently with the truth of their own view. Here's how they might do that. As I pointed out in Chapter One, there are ways of understanding talk of agent's responding to facts where the sort of response in question is not the kind of response we're interested in, that is: a rational response, one

which is an operation of the capacity to ϕ for reasons. Moreover, this goes even if the fact to which the agent is responding, in the non-rational way at issue, just so happens to be a normative reason. The chicken sexer believes that the chick is male *in response to* the fact, sub-personally detected by him, that the chick is letting off a distinctively male pheromone. That would not be a response to a reason in the sense we're interested in, even though the fact in question happens to be a reason for the corresponding belief. Hughes' hiker case does involve a non-rationalising causal link between an attitude and a fact which happens to be a reason for it. So the proponent of the Epistemic Thesis can say that Hughes finds it intuitive that the Hiker feels relieved because there's a hut in the distance in the rationalising sense of 'because' because he has conflated the rationalising sense with the more brutally causal sense just intimated. Thus, Hughes' intuition is explained away consistently with the truth of the Epistemic Thesis.

The problem with this initial reaction to Hughes' case is that it is dialectically ineffective. For let's suppose Hughes were simply to say: 'no, it really is the case that Henry feels relieved because there's a hut in the distance, in the rationalising sense'. What could the proponent of the Epistemic Thesis say then? Well, what they would have to say is that the onus of proof is on Hughes to demonstrate that there is anything more than the causal sense of 'because' in operation. Now, this response essentially takes it to be the case that the Epistemic Thesis is well-motivated and hence that the onus of proof is on Hughes to demonstrate that his verdicts are correct. However, Hughes' verdicts about his case serve to undermine the very motivation for the Epistemic Thesis. That's because the Epistemic Thesis is motivated by appeal to an inference to the best explanation concerning the Hornsby-Hyman cases. The best explanation of why the subject fails to ϕ because p, so the inference goes, is that the subject doesn't know that p. But if Hughes' verdicts are correct, then there is an alternative explanation available: the subjects fail to ϕ because p because their beliefs that p are not caused by p. Moreover, if Hughes is right then that explanation would be *better* than the explanation in terms of the agent's lack of knowledge because it is more general: it is able to explain why subjects are not in a position to respond to reasons in a greater range of cases than the hypothesis that responding to reasons requires knowledge, given that the causal condition is weaker than knowledge. So the obvious response to Hughes will not work, as it stands.

I think, however, that the obvious response to Hughes can be rescued, because it can be shown, independently of the debate concerning the truth of the Epistemic Thesis, that the hypothesis that responding to normative reasons doesn't require one's belief that p to be caused by p. Consider the following case:

On his day off, Billy decides to go and read his novel in a coffee house somewhere in town. He knows that there are a lot of coffee houses on Main Street. So: he catches the bus to Main Street for the reason that there are a lot of coffee houses there.

The case of Billy and his day off is intended to be an entirely ordinary case of ϕ -ing because p: Billy does indeed catch the bus because there are a lot of coffee houses on Main Street in the rationalising sense of ‘because’. If the hypothesis that responding to normative reasons requires a causal link between one’s belief and the fact which constitutes one’s normative reason is right, the Billy’s belief will have to count as being caused by the fact that there are a lot of coffee houses on Main Street. But it is implausible that there is such a causal relation between his belief and that fact. His belief might be caused by an inference from a number of other beliefs, some of which record the presence of a number of different coffee houses on Main Street, but the fact that there are a lot of coffee houses there is not amongst the causes of his belief – what would it be for such a quantificational fact to cause his belief? It isn’t true, in general, then, that responding to reasons requires a belief which is caused by the fact that is one’s reason.

What is the upshot of this? The upshot, I think, is that Hughes’ verdicts no longer serve to undermine the motivation for the Epistemic Thesis. For central to the thought that they undermine the motivation for it is the availability of an alternative hypothesis: that ϕ -ing because p, in the rationalising sense, requires believing that p in a way that’s caused by p. But that hypothesis has been shown to be independently false, and so isn’t available as an alternative with respect to explaining what’s going on in the Hornsby-Hyman cases. It follows in turn that those cases can be relied upon in order to motivate the Epistemic Thesis,⁶ and this serves to shift the onus onto opponents of it like Hughes to prove that what’s going on in cases in which the subject doesn’t know that p but does ϕ because p is that we have something more than a merely causal ‘because’ in operation. Until such proof is forthcoming, we can conclude that the Epistemic Thesis is true.

I am not quite finished with my initial exploration of what responding to a normative reason requires. I want, finally, to argue for the truth of a condition on responding to normative reasons which constitutes something of an extension of Hornsby and Hyman’s Epistemic Thesis.

I’ve been defending a certain thesis linking awareness of one’s normative reasons with one’s being able to ϕ in response to those reasons. But being aware of one’s reasons in such cases is not the only psychological status that is supposed to enable one to ϕ in response to the facts. As I’ve already noted, the subject’s *taking p to have the normative significance it does* is also a condition required for responding to normative reasons.

I do not want to offer an account of this taking-to-be-normative condition in this essay. But what I want to point out here is that, whatever that condition turns out to be, we can make sense of the thought that the subject *knowledgeably* takes p to have the normative significance it does. This will be a case of the subject not only getting it right about the

⁶Indeed, the case of Billy could easily be converted into a fresh motivation for the Epistemic Thesis. All we’d have to do is contrast the case of Billy with a corresponding bad case in which he is told by an ordinarily reliable informant that Main Street boasts many coffee houses, where it happens to be the case that that’s so, but where the informant doesn’t have an opinion on it either way.

normative status of the relevant fact, but getting it right in a non-accidental way that is at least analogous to the non-accidental way in which one ends up believing truly in cases of knowledge. Of course, this point is easy to accommodate if taking p to be a normative reason just is believing that p is a reason in favour of one's ϕ -ing. But even if we remain neutral on that issue, still it must be agreed that there is some sense in which the condition can be exemplified knowledgeably by the subject.

Moreover, we have an intuitive grip on what that would amount to: knowledgeably taking p to be a reason for one to ϕ would be knowing it to have a certain normative status so that one knows how to weigh it up properly against countervailing reasons and how to weigh it together with other considerations on the same side. But with that point in tow, I suggest that it is at least initially plausible that one needs to knowledgeably take p to be a reason for one to ϕ if one is to count as ϕ -ing in response to the normative reason that p , in the sense we're interested in. Responding to the reason that p is a matter of responding to p *qua* reason. This suggests that responding to the reason that p in the sense at issue involves an appreciation of the normative status of p . I suggest that it is quite natural to cash this out in terms of knowledgeably taking p to have the normative status it does.

This section has explored the issue of what it is to ϕ in response to the normative reason that p . I have argued for the following necessary conditions on responding to normative reasons: (i) that one's ϕ -ing is subject to a rationalising explanation which involves one's reason; (ii) that p is a normative reason for one to ϕ , so that p is true and p favours one's ϕ -ing; (iii) one knows that p ; (iv) one knowledgeably takes p to be a reason in favour of one's ϕ -ing. Given (iii) and (iv) we can say that when one ϕ s in response to the normative reason that p one's ϕ -ing manifests one's knowledge of one's reason *qua* reason. (iii) and (iv) should be taken to be developments of (DT) and (NT), described in Chapter One: responding to the reason that p doesn't just require believing that p and taking p to be a reason, but knowing that p and knowing it to be a reason.

5.2 Williams' Dictum

Having gotten somewhat clearer on what is involved in the success condition I now turn to Williams' claim that the form of explanation associated with ϕ -ing for reasons is unaffected by the true/false distinction. This section offers an interpretation of that claim, which I believe to be at the core of Reductionist thinking about the success condition.

Let us return to Smith and his penchant for free coffee. Smith's action is subject to a rationalising explanation. Since Smith is in the good case, we can express the, or a, rationalising explanation to which his action is subject by saying: 'Smith walks to the quad because there is free coffee available there'. Since Smith walks to the quad for a reason, he must believe the proposition which constitutes his reason: Smith believes that there is free coffee available in the quad. We can also provide a rationalising explanation

of Smith's action which makes explicit reference to Smith's belief, if we like. We can do so by saying: 'Smith walks to the quad because he believes that there is free coffee available there'. We can offer a rationalising explanation using the latter form of words even if Smith is wrong about whether there is free coffee available.

Williams suggests that whether or not Smith's belief that there is free coffee available in the quad is true, the *form* or *type* of explanation to which Smith's action is subject must remain the same. What does this mean? We can immediately rule out some initial suggestions. Williams is not saying that the *logical form* of the explanation is the same whether the belief is true or false. I know of no reason to deny that, and in any case Williams seems to have in mind something more substantial. Williams is also not saying that whether or not Smith's belief is true there is a *rationalising* explanation of Smith's action available. That much is granted by everyone: rationalising explanations apply in both good cases and bad cases, and Williams is not asserting something which is a platitude in this area. So what is he saying?

Consider the following set of efficient-causal explanations:

- (i) The tree collapsed because it was hit by a strong wind.
- (ii) The left side of the man's face is numb because the dentist administered local anaesthetic to it.
- (iii) The tree collapsed because it was struck by lightning.
- (iv) The man cannot talk clearly because the dentist administered local anaesthetic to the left side of his face.

Operating at the intuitive level, there is a clear sense in which we have four *different* explanations here. What makes them different explanations is that none of them involve the same pair of *explananda* and *explanantia*, even if some of them ((i) and (iii)) share an *explanandum* and some of them ((ii) and (iv)) share an *explanans*. But there is also an intuitive sense in which they are of the same *form* or *type*. All of the above explanations belong to the type *efficient-causal explanation why*. They make the relevant *explanandum* intelligible by making it intelligible *why*, as opposed to how, when or where, it is the case and by dint of appealing to a causal link between the entities corresponding to *explanandum* and *explanans*.⁷

Now consider the following pair of constitutive explanations why:

- (v) The wall is hard because it is made out of brick.

⁷Strictly speaking, we can distinguish between *two* types under which the explanations (i)-(iv) fall: *explanation why* and *causal explanation why*, where the latter is more specific than the former. In order to keep things simple I talk of *the* type to which an explanation belongs. That talk is really shorthand for talk of *the most specific type* to which the explanation belongs. I leave open, despite the way I talk, the possibility of there being more than one type under which an explanation falls.

(vi) The statue is losing its beauty because the clay from which it is made is crumbling.

Again, there is a clear sense in which each of these explanation are *different* explanations: they have different *explananda* and *explanantia*, after all. But there is also a clear sense in which they are of the same type: they are *constitutive explanations why*. Like the form associated with (i)-(iv), they make the relevant *explanandum* intelligible in the sense of making it intelligible *why* it is the case. But unlike the form associated with (i)-(iv), (v)-(vi) make the relevant *explanandum* intelligible by dint of appealing to a constitutive link between the entities corresponding to the *explanandum* and *explanans*.

Finally, consider the following pair of teleological explanations why, which I assume for the sake of the discussion here cannot be reduced to efficient-causal or constitutive explanations:

(vii) The plant grows towards the window in order to catch the light.

(viii) The spider weaves the web in order to catch its prey.

Once again, there is a clear sense in which these explanations are different explanations but are of the same form: they are *teleological explanations why*. Like the forms associated with each of our other set of explanations, they are explanations *why*. But unlike the members of each of the other two sets they make the relevant *explanandum* intelligible by appealing to the fact that the occurrence which corresponds to the *explanandum* in some sense contains or has the goal specified by the *explanans*.

For each set of explanations we can say this of its members: they are different explanations but they have the same form. It is by dint of belonging to the relevant form that each distinct explanation in the set has certain properties that it shares with the others, properties which are essential to the form or type in question. It is because each of (i)-(iv) belong to the form they do that they require for their truth the obtaining of a causal relation; it is because (v)-(vi) belong to the form they do that they require the obtaining of a constitution relation; and it is because (vii)-(viii) have the form they do that they require the exemplification of a having-a-goal property by the relevant event or process. Moreover, it is because all of the explanations are explanations *why*, an aspect shared by each of the explanatory forms examined, that they must have true *explananda* and, in the case of the causal and constitutive relations, true *explanantia*.⁸

We can think of a form of explanation as a *way of making something intelligible*. One way of making the fact that the tree collapsed intelligible is to make it intelligible *why it collapsed by appeal to the cause*. This differs from making it intelligible *how* it collapsed, or *when* it collapsed. But it also differs, as we have seen, from making it intelligible *why it collapsed by appeal to its constitution* and *why it collapsed by appeal to an end it*

⁸I am not suggesting here that the exemplification of the form explains why the explanations have the properties in question. I leave it open that having the properties in question is constitutive of having the form.

had. There are ways and ways of making something intelligible. The different forms of explanation correspond to these different ways of making things intelligible. The particular explanations which fall under the types all make things intelligible in the same sort of way, they just make different *explananda* intelligible, by appeal to different *explanantia*.

Forms of explanation are generic: they do not require, of any given exemplar, that it has a particular *explanans* or *explanandum*. Nevertheless, forms of explanation require that their exemplars have a certain *kind* of *explanans* and *explanandum*. For example, all explanations why require that their *explanantia* and *explananda* are true, if they are capable of truth and falsehood at all, and causal explanations why require that their *explanantia* and *explananda* correspond to, or are, factors that stand in a causal relation to each other. To think of a form of explanation as requiring *explananda* and *explanantia* of certain types is to think that it is essential to the way that the explanation makes things intelligible that an *explananda* or *explanantia* of that type is on the scene, even if the particular *explanans* or *explanandum* appealed to is not required by the very form of intelligibility at issue.

We can think of rationalising explanation as a form of explanation. *Rationalising explanation why* is a distinctive way of making things intelligible. (i)-(viii) are not rationalising explanations.⁹ Rationalising explanations are a form of explanation why, and so have that in common with the explanations just considered, but we are to distinguish rationalising explanations from the sort of efficient-causal explanations, constitutive explanations and teleological explanations just considered nevertheless.

It is a good question what the form *rationalising explanation* amounts to. One thing we can say is that for an explanation to be a rationalising explanation requires that it takes a fact which records the obtaining or occurrence of a judgement sensitive phenomenon as *explanandum* and appeals to some factor as *explanans* which involves what the subject takes to be a reason in favour of their engaging in the judgement sensitive activity in question. That's part of what the form of intelligibility involved in rationalising explanation requires. I will be adding to this conception of rationalising explanation as we go along.

With a rough characterisation of the notion of a form of explanation in play, we are now in a position to understand Williams' claim. When Williams says that the form of explanation to which Smith's action is subject is not affected by whether Smith is believing truly or falsely the form that he is focusing on is the form *rationalising explanation*: explanations why given in terms of what the agent takes to be the reasons in favour of their ϕ -ing. What he is saying is that that form is not affected by whether Smith's belief is true. As I have said, I do not think we should interpret this as the suggestion that whether Smith's action is subject to a rationalising explanation at all fails to depend on the cor-

⁹This is not to deny that rationalising explanations are efficient-causal, nor that they are teleological. The thought is that the most specific type to which rationalising explanations belong differs from the most specific types to which (i)-(viii) belong. Consistently with that, rationalising explanation might still exemplify, as a more generic type, the efficient-causal and teleological forms exemplified as the most specific types by (i)-(iv) and (vii)-(viii) respectively.

rectness of Smith's belief. Rather, Williams is agreeing that Smith's action is subject to a rationalising explanation across both cases, but he is intending to make a comment about the nature of that way of making things intelligible.

What I want to suggest is that Williams' claim can be read in two ways. On the one hand, Williams could be read as saying that necessarily, for every explanation which exemplifies the form *rationalising explanation* – a form of explanation the availability of which depends on the agent taking p to be a reason in favour of their ϕ -ing – *that very explanation* could be true if the subject's belief is true or if it's false. The truth of the *particular explanation* doesn't require the subject's belief to have a particular truth-value. All particular rationalising explanations are unaffected by the true/false distinction.

On the other hand, Williams could be read as saying that whenever an agent's ϕ -ing is subject to a rationalising explanation, it cannot be that the *form of explanation* which it exemplifies requires the subject's belief to be true. It cannot be that there is a way of making the subject's ϕ -ing rationally intelligible which requires that the subject gets their reasons right. As we might put it: whenever an agent's ϕ -ing is subject to a rationalising explanation *the way in which the judgement sensitive phenomenon in question is made intelligible* doesn't require that the agent's belief is true or false. There is one type of thing being a rationalising explanation is, and it is neutral on whether or not p is true. On this reading, Williams is denying that there can be an explanation which is rationalising but in a way that requires that p has a particular truth-value: there is no such thing as the type *rationalising-explanation-in-terms-of-a-normative-reason*.

These two readings are genuinely distinct. To say that the form to which all rationalising explanations belong is such that the truth of the explanation doesn't require that the subject's belief has a particular truth-value, is consistent with the thought that the truth of the particular explanation in question does require that the belief has a particular truth-value. It's just that, if it does, that will not be by dint of its exemplifying the form it does. So Williams' claim on the second reading does not entail the claim on the first. However, the claim read in the first way does entail the claim read in the second. For if no rationalising explanation requires the truth of what the subject believes, then it cannot be that are any rationalising explanations which do have that requirement by virtue of their form.

There is a way of generalising Williams' claim and it is the claim in this generalised form which I want to focus on. The way to generalise it is to have it focus not just on whether p is true or false but on whether p is a normative reason for S to ϕ , so that the claim is that rationalising explanation is independent of whether p is a genuine reason for S to ϕ . That claim is more general than Williams' original because if rationalising explanation doesn't require p to be a genuine reason for S to ϕ then it cannot require p to be true, for p must be a fact if it is to be a genuine reason.

Generalised in that way Williams' claim is still ambiguous. First, it can be read as the claim that all particular rationalising explanations do not require that p is a normative

reason for S to ϕ in order to be true; second, it can be read as the claim that the type to which all rationalising explanations belong doesn't require p to be a normative reason for S to ϕ . The claim read in the first way is stronger than the claim read in the second: the former entails the latter but not *vice-versa*. Thus, I will label the first *Williams' Dictum_s* and the second *Williams' Dictum_w*. Let us codify the two theses here:

Williams' Dictum_s Necessarily, for any rationalising explanation, *RE*, *RE* can be true whether or not p is a normative reason for S to ϕ .

Williams' Dictum_w Necessarily, for any rationalising explanation, *RE*, *RE* does not make the agent's ϕ -ing intelligible in a way that requires p to be a normative reason for S to ϕ .

In the next section I will be examining what Davidson and Dancy could say about the success condition and in so doing I will provide some illustrative examples of the distinction between each of Williams' Dicta. But before moving on to that, some initial clarificatory comments are in order. The distinction between the two theses hinges on the distinction between a particular explanation and the form exemplified by it. The thought is that an explanation might exemplify a certain requirement by dint of its form, or it might exemplify a certain requirement, but not by dint of its form. This seems to me to be a perfectly good distinction to draw, as our investigation of the three sets of non-rationalising explanations above can be relied upon to show. Take, for example, explanation (i). The form of the explanation, that is: the way it makes its *explanandum* intelligible, is *causal explanation why*. The truth of (i) requires that the particular fact to which it appeals – that the tree was hit by a strong wind – obtains. The explanation would be false if that fact didn't obtain. But the form of (i) does not require that fact to obtain: it is not built into the general way in which things are made intelligible by (i) that the particular fact about the wind obtains. So an explanation might exemplify a requirement even though it does not do so by dint of its form.

This should help us understand the distinction between Williams' Dictum_s and Williams' Dictum_w. The requirement at issue is that the explanation is true only if p is a normative reason for S to ϕ . The former thesis says that this is not a requirement of any particular rationalising explanation. The latter says that it is not a requirement built into the form exemplified by any particular rationalising explanation. As should be clear, the latter might be true whilst the former false.

5.3 Reductionism

I have advertised the main target in this chapter as Reductionism. But I have not talked about Reductionism so far. This section argues that Reductionism is logically equivalent

to Williams Dictum_w: it is the picture of the good case delivered to us by that thesis. It follows that undermining Williams' Dictum_w is a way of undermining Reductionism. That is the strategy to be pursued later on.

I want to begin by returning to the competing conceptions of rationalising explanation offered by Davidson and Dancy, each of which constitute versions of Reductionism. I want to show that each account can be developed so as to accommodate the success conditions discussed in §5.1. One point which will emerge is that although Davidson and Dancy must accept Williams' Dictum_w, they needn't accept Williams' Dictum_s. After that, we'll be in a good position to appreciate precisely what Reductionism about the success case amounts to and how it is connected to Williams' Dicta. It should be noted before I continue, however, that what I offer in §§5.3.1-2 is not an interpretation of what Davidson and Dancy actually commit to. Rather, what I offer here is a sketch of how they *could* develop their theories in order to accommodate the success case.

5.3.1 Success for Davidson

According to Davidson, we're to think of the type to which rationalising explanations belong as neutral on whether *p* is a genuine reason for the subject to ϕ : in so far as an explanation is of that type, it doesn't follow that *p* is a fact which favours their ϕ -ing. But Davidson thinks of the rationalising type in a distinctive way. According to Davidson, an explanation's being a rationalising explanation requires it to take the agent's ϕ -ing as *explanandum* and the state of *S*'s believing that *p* as *explanans*, where the explanation is underwritten by an efficient-causal relation holding between the state and the ϕ -ing. An explanation of that type holds in both the good case and the bad case.

Now let us ask Davidson the following question: what is it to ϕ in response to the normative reason that *p*? Since that condition constitutively involves the obtaining of a rationalising explanation, Davidson will say that part of the answer is that ϕ -ing in response to a normative reason involves the agent ϕ -ing because of their belief that *p*, where the explanatory link is underwritten by a causal relation between the two. But Davidson cannot rest content with that answer, because ϕ -ing in response to a normative reason requires that *p* is true and is a normative reason, but ϕ -ing because of one's belief that *p* doesn't entail that: the type of explanation present doesn't guarantee that those features of the success case obtain. So he will have to add: ϕ -ing in response to the normative reason that *p* consists in ϕ -ing because of one's *true* belief that *p* – where the explanation is causal and *p* favours one's ϕ -ing. But what about the Epistemic Thesis: that *knowing* that *p* is required for ϕ -ing in response to a normative reason?¹⁰ Davidson can incorporate this point just by dint of the fact that one's knowledge in the good case is one's belief. Thus, properly developed, Davidson's position on the good case is this: to ϕ in response to the reason that

¹⁰I ignore the fourth condition identified in §5.1 in order to ease the discussion.

p is to ϕ because of one's knowledge that p, where p favours one's ϕ -ing and where the explanation is efficient-causal.

At this point, what the Davidsonian position amounts to exactly depends on how the state of knowing is conceived. If, on the one hand, the state of knowing were thought of as reducing to the state of believing, plus truth and warrant,¹¹ Davidson would end up saying that the *explanans* in the good case is identical to the *explanans* in the bad, and hence that the two explanations would be the same explanation. The *explanans* would be the same because the state the subject is in in the bad case would be token identical to the state they are in in the good case. It's just that the *explanans* in the good case would happen to have certain properties which give it the status of knowledge. If, on the other hand, the state of knowing were thought of as a primitive type of propositional attitude in its own right, as Williamson (2000) would have us believe, then Davidson would be committed to saying that the *explanans* in the good case is genuinely different to the *explanans* in the bad. For we would have a different kind of state present in the good case than in the bad. Either way, it's worth emphasising, the *type* of explanation would be the same.¹²

So Davidson's position needs to be developed so that he incorporates the point that knowing is required for responding to a normative reason. This point is incorporated by an identification of the state of believing with the state of knowing in the good case. This position then divides into two sub-positions, each corresponding to the two different ways of conceiving of knowing. The first position, which utilises the standard conception of knowledge as a special kind of believing, commits Davidson to the claim that particular rationalising explanations are exactly the same in the good case and bad: both appeal to a state of believing. The second position, which utilises the Williamsonian conception of knowledge, generates the result that the particular explanations are different, for they appeal to different kinds of state. Both explanations are of the same type: rationalising explanation, thought of as a type which is neutral on whether the subject is in the good case.

There are two points to highlight about the ways of developing the Davidsonian position which have just been explored. First, each way of developing the position differs with respect to its relationship with Williams' Dicta. The second way of developing the posi-

¹¹ Here I identify warrant with whatever factor converts true belief into knowledge.

¹² The initial way of incorporating the point that responding to a normative reason requires knowing was to say that the *explanans* of the good-case rationalising explanation *is* the state of knowing. The 'is' there can be an 'is' of identity, if the standard conception of knowledge is assumed. If the Williamsonian view is assumed then the 'is' could still be the 'is' of identity, in which case a disjunctive conception of believing, of the sort rejected by Williamson himself, will have to be operated with. That is the way I have tacitly assumed things would go. However, it could be interpreted as an 'is' of constitution if the Williamsonian conception of knowledge is assumed too. This raises the possibility that Davidson could incorporate the epistemic point in a different way, in order to avoid committing himself to the claim that the explanations are different across good cases and bad. What he could say is that the *explanans* is a belief in both cases, it's just that in the good case the belief is constituted by a state of knowing. That would be a third way of developing Davidson's position which I leave out of the text in order to not make the discussion even more complicated than it already is.

tion combines a rejection of Williams' Dictum_s with an acceptance of Williams' Dictum_w. It accepts the latter because it says that rationalising explanations do not belong to a type which require that p is a normative reason for S to ϕ . But it rejects the former because it says that explanations in the good case take as *explanantia* states of knowing, which just are the states of believing that cause one's ϕ -ing. So the particular explanation to which one's ϕ -ing is subject in the good case is not identical to the explanation to which one's ϕ -ing is subject in the bad case. On the first way of developing the position, however, both of Williams' Dicta are maintained. Williams' Dictum_w is maintained in the same way and Williams' Dictum_s is also accepted because the *explanans* is a state of mind identical across the good case and bad.

Davidson's position thinks of the form exemplified by all rationalising explanations as neutral, so that if a particular rationalising explanation takes as its *explanans* an item which can function as *explanans* only in the good case – for example, the fact that p or the state of knowing that p – then that will be accidental from the point of view of the way in which it makes the relevant *explanandum* intelligible – it will be accidental from the point of view of the form of the explanation. This brings me to my second point. Although the state of knowing can figure in Davidsonian rationalising explanations, as it does on each development of the basic position, it cannot be that the agent's ϕ -ing in the good case is explained by a state of knowing *qua* knowing. The way in which rationalising explanations make ϕ -ings intelligible, for Davidson, is simply by dint of connecting the ϕ -ings with the relevant causally active state of mind that constitute the *appearance* of normative reasons to the subject, not to those states of mind which constitute the subject's *awareness* of such reasons. So even though knowing can figure in rationalising explanations in the good case, for Davidson, it is accidental to the way in which the explanation makes the relevant judgement sensitive phenomenon intelligible that the state of knowing plays the role it does. There might be a sense in which knowing does explanatory work for Davidson, but we must not be misled into thinking that it is *qua* knowing that the agent ϕ s.

5.3.2 Success for Dancy

We've looked at how the Davidsonian position might cope with the good case and I've put on the map two ways of developing the theory so as to cope with it. On each way of coping with it, the *explanans* in the good case is a state of knowing, but only on one way of developing it does that get us the result that Williams' Dictum_s is true. But now let's move on to examining Dancy's Reductionist position. We'll find, again, that there are two different ways of developing Dancy's position which differ with respect to their relationship with Williams' Dicta.

According to Dancy, we're to think of the type exemplified by all rationalising explanations as neutral with respect to whether p is a normative reason for S to ϕ . The way in

which things are made intelligible by rationalising explanations does not essentially involve reference to a normative reason. The type of explanation is to be thought of as an explanatory link between the agent's ϕ -ing and what they take to be a normative reason: p . The explanation is underwritten by a primitive type of non-causal relation which links p with the agent's ϕ -ing. The sort of relation holds even when p is false. This makes the explanation non-factive: it can be true even if its *explanans* is false, or does not obtain. The sort of explanation here is enabled by the subject believing that p and taking p to be a normative reason for them to ϕ , even though neither of those conditions are part of the *explanans* of the rationalising explanation.

Now let's ask the following question: how should Dancy account for the success condition of ϕ -ing in response to the normative reason that p ? First, Dancy will say that it involves the obtaining of an explanation that falls under the type *rationalising explanation*. But since Dancy conceives of that type as failing to suffice for the subject to be in the good case, his answer cannot stop there. He will have to say, in addition, that it involves p being true and being a genuine reason in favour of the agent's ϕ -ing. He will also have to acknowledge the point that the success condition requires knowing that p . How should he incorporate all of this into his account?

I suggest that the basic idea should be that the type of explanation which holds in the good case doesn't require p to be true, or be a normative reason or that one knows that p . But the particular explanation does take the fact that p as *explanans*. Dancy can then incorporate the epistemic point by saying that the belief which enables the explanation is a state of knowing. So in the good case, the explanation which holds explains why the agent ϕ s by appeal to a fact which favours the agent's ϕ -ing, and the explanation is enabled by a piece of knowledge on the agent's part.

However, there are two ways of developing this conception of the good case which give us different results. The variation depends on how Dancy thinks of facts. If, on the one hand, he thinks of facts as entities which exist across both good cases and bad, but which only obtain – only are facts – in the good case, then he can agree that the *explanans* in the good case is identical to the *explanans* in the bad case. If, on the other hand, Dancy thinks that the fact that p does not exist in the bad case, so that whatever facts are they are facts *essentially*, then he is committed to thinking of the *explanans* of rationalising explanations in the good case as distinct from the *explanans* in bad cases and hence as the particular explanations as being different. In other words: if Dancy thinks that the identity of a fact can be held fixed whilst varying whether or not it obtains, then Dancy will be committed to saying that the particular explanation in the good case has the same *explanans* as in the bad case, and hence that the particular explanations are the same. If, on the other hand, Dancy were to say that the obtaining of a fact is an essential feature of it, then he will have to say that the *explanans* in the good case is not the same as that in the bad case, and hence

that the two explanations are different.^{13,14}

Again, there are two comments which need to be made about these possible developments of Dancy's basic position. First, they differ with respect to their relationship with Williams' Dicta. The second way of developing Dancy's basic idea is committed to Williams' Dictum_w but rejects Williams' Dictum_s. It's committed to the former because it says that the form of explanation in each case is the same – a form which is neutral on whether *p* is a reason for the subject to ϕ . It rejects the latter because it says that the explanation in the good case has a different *explanans* to the corresponding explanation in the bad – an essentially obtaining fact in the former, and a distinct entity, or perhaps no entity at all, in the latter. On the first way of developing the position, however, according to which the *explanans* is a fact which obtains in the good case but the very same entity, just non-obtaining, in the bad, the *explanantia* are identical and so the particular explanations are identical too. Thus, both of Williams' Dicta are accepted.

The second point to highlight is that, even though on both positions the normative reason that *p* thought of as a fact with a distinctive normative profile is the *explanans* of the rationalising explanation in the good case, it cannot be that the normative reason explains *qua* fact or *qua* its normative status: the status of the normative reason as a fact is not an explanatorily relevant feature of the *explanans* in the good case, and neither is its normative character. Again, this is because the very way in which the explanation makes the agent's ϕ -ing intelligible makes no appeal to the status of *p* as a fact or as a normatively significant item. The form of explanation ignores those properties of the *explanans* and hence those properties don't do explanatory work for Dancy, however the position is developed in order to accommodate the good case. Rationalisers get to be rationalisers for Dancy just by dint of being *apparent* normative reasons, not by dint of being the genuine article. That's so even if the genuine article is what happens to do the explaining in the good case.

5.3.3 Reductionism & Williams' Dicta

We've seen, then, that there is flexibility with respect to how the good case is understood by Davidson and by Dancy. Both can agree that a normative reason-involving condition can explain why the agent ϕ s in the good case: a state of knowing, for Davidson, the normative reason that *p* itself, for Dancy. Whether this requires the proponent of the theory to reject Williams' Dictum_s depends on how the suggestion is further developed. But both agree that the type of explanation is neutral on whether the subject is in the good case and hence

¹³This metaphysical issue mirrors the metaphysical issue of whether we should think of knowing as consisting in a state that can exist without enjoying the status of knowledge or whether we should conceive of it as a state which is essentially knowledge. That issue in the metaphysics of mind is what the difference between the two versions of the basic development of Davidson's position turns on.

¹⁴Dancy's preferred position on this matter, at least in *Practical Reality*, is that facts, and hence normative reasons, are states of affairs which obtain. States of affairs are thought of as items which can exist without obtaining (Dancy, 2000:114-118). His preferred view would thus be a variant of the first development of his position.

are committed to saying that it is not *qua* knowledge or *qua* normative reason that the relevant *explanans* does its explanatory work. That aspect of the Williamsian picture must be upheld.

I now wish to introduce my main target in this chapter: Reductionism about the condition of ϕ -ing in response to the normative reason that *p*. The basic idea of Reductionism is this. We are to think of responding to the normative reason that *p* as consisting in the holding of a rationalising explanation. But the Reductionist says that we are to think of the type exemplified by all rationalising explanations as neutral: in so far as the rationalising explanation belongs to the type it does, it fails to follow that *p* is a genuine reason for *S* to ϕ . Making an agent's ϕ -ing intelligible in the distinctively rationalising way can never consist in making it intelligible by appeal to a normative reason, or else to the state of mind which constitutes one's awareness of a normative reason. Now, that an explanation of that type holds cannot suffice for the subject to count as responding to a normative reason, for an explanation of that type holds in the bad case. So it must be that responding to a normative reason cannot simply consist in the holding of an explanation of that type. It must instead consist in an explanation of that type holding, but, in addition, a set of other factors holding which are extra to the type of explanation in question. These factors will include the truth of *p* and its normative status, as well as one's knowledge of *p* and its normative status. The extra factors will, as we have seen, been assigned an explanatory role: they will either be the *explanans* of the particular explanation, be a part of that *explanans*, or be an enabling condition for particular explanation.

In saying that the relevant factors are *extra* to the type of explanation in question I intend to say two things. First, that they are not guaranteed to obtain just because there is an explanation of that type on the scene. The truth of *p*, its normative status, one's knowledge of *p* and its normative status, and whatever other factors are involved in the good case other than the type of explanation, are not guaranteed to be there just by dint of there being an explanation of that type on the scene. Second, I intend to say that even if those factors enter into the explanation in the ways at issue, the explanation doesn't appeal to them *qua* the conditions that they are: it is not *qua* true, or *qua* normative reason, or *qua* knowledge that the relevant factors explain, if they do.

Reductionism is logically equivalent to Williams' Dictum_w. First, Reductionism entails Williams' Dictum_w simply because it has a commitment to that thesis built into it. The Reductionist precisely thinks that the type exemplified by all rationalising explanations is neutral on the presence of normative reasons. But second, Williams' Dictum_w entails Reductionism. That's because if one subscribes to Williams' Dictum_w then one will have to say that ϕ -ing in response to a normative reason constitutively involves a rationalising explanation, and since the type to which it belongs is neutral one will have to say that, in addition to there being an explanation of that type present, also there are some extra-explanatory factors present which guarantee, together with the explanation, success.

So Reductionism just is the picture of the good case one ends up with if one subscribes to Williams' Dictum_w and if one subscribes to that picture then one must subscribe to Williams' Dictum_w. Refuting the latter, then, constitutes a refutation of Reductionism.

I now want to move on to present the conception of the good case I want to defend, which involves a commitment to a thesis inconsistent with Williams' Dictum_w and hence inconsistent with Reductionism. I will be arguing for that thesis later, and that is how my refutation of Reductionism will work.

5.4 The Essential Normativity Thesis

My preferred way of rejecting Williams' Dictum_w can be brought out by considering the following passage from Hyman's *Acting for Reasons: Reply to Dancy*:

As we have seen, we have the idea of reasoning from a premise, which may be known, or merely believed. . . and which may be true or false. But as well as the idea of reasoning from a premise, we also have the idea of being guided by a fact. As noted earlier, if someone is said to have been guided by a certain fact, this means that he took it into consideration when he modified his thought or behaviour in some way, or decided what to think or what to do. This, I suggest, is how we should understand the idea of a fact being a person's reason. (Hyman, 2011:355-356)

In this paper at least, Hyman accepts, following Dancy and other proponents of the Motivating Reasons Postulate, that the agent ϕ s for the reason that p in both the good case and the bad case.¹⁵ With that in the background, we can understand what Hyman is doing here in the following way. Hyman wants to draw a distinction between two kinds of relations that might hold between the reason for which the subject ϕ s and their ϕ -ing, when the subject ϕ s for the reason that p. The first relation he labels 'reasoning from a premise'. Reasoning from a premise is supposed to be a relation which holds whether or not p is a *bona fide* normative reason for the subject to ϕ . The obtaining of that relation suffices for the agent to ϕ for the reason that p, as Hyman conceives of the latter condition.¹⁶ But in addition to that normatively neutral relation, there is also the relation he labels 'being guided by a fact'. We can think of this second relation as a relation that is identical to what

¹⁵Hyman's relationship with the Motivating Reasons Postulate is unclear. In *How Knowledge Works*, *The Road to Larissa*, and *Knowledge and Evidence* he rejects the claim that S ϕ s for the reason that p in the bad case, in which case in those papers we should rank him alongside Raz, Alvarez, Roessler and Stout in rejecting the Postulate. But in *Acting in Ignorance*, as we have seen, he sides with Dancy and the rest in accepting it, without flagging it up that he's changed his mind.

¹⁶Concerning the relation of reasoning from a premise, Hyman points out that we'll have to be careful to distinguish it from the relation that holds between the subject's ϕ -ing and p, when p is merely something that has been *supposed* to be true, as opposed to *believed* to be true, during the course of reasoning which led up to the subject's ϕ -ing.

I've been calling 'responding to the normative reason that p', so that the relation holds between the subject's ϕ -ing and their reason only in the good case. Being guided by the fact that p is a relation which requires p to be a normative reason, and requires one to know that p. Treating p as a premise requires neither of those things.

Importantly for my purposes, Hyman thinks that the relation of *being guided by* and the relation of *treating as a premise* are explanatory relations. By that I mean that Hyman intends the relations to be such that if they obtain, then the agent's ϕ -ing is subject to a certain explanation why, indeed a *rationalising* explanation why. But part of the purpose of separating the two relations is to make the stronger point that the corresponding *forms* of rationalising explanation are themselves different. On the one hand, there is the form of rationalising explanation which involves an explanatory link between the agent's ϕ -ing and the known fact that p, which is made available by the obtaining of the *being guided by* relation. On the other hand, there is the form of rationalising which involves an explanatory link between the agent's ϕ -ing and their belief that p, which is made available by the obtaining of the *treating as a premise* relation.

The latter sort of rationalising explanation is neutral on whether p is a normative reason for S to ϕ and hence is a type which can obtain in both good cases and bad. It is a type of explanation which requires, of the explanation which exemplifies it, a judgement sensitive phenomenon – or else a fact which records the presence of one – as *explanandum* and the state or fact of the subject's belief that p as *explanans*. It also requires the obtaining of the 'treating as a premise' relation. The former type of rationalising explanation essentially appeals to the fact that p as *explanans*. It requires, of the explanation which exemplifies it, that it takes a judgement sensitive phenomenon – or the fact of one – as *explanandum* and the fact that p as *explanans*. It also requires obtaining of the 'being guided by' relation as well as requiring that the subject knows that p, for knowing that p enables the 'being guided by' relation to obtain. It follows, given that the form of rationalising explanation which corresponds to the 'being guided by' relation has this character, that it is a form of rationalising explanation which is available only if the subject is responding to a normative reason. That's because no explanation available in the bad case essentially involves reference to what can be said in favour of the subject's ϕ -ing, even if it is a bad case in which p happens to be a normative reason.

It should be clear that Hyman's position involves the rejection of Williams' Dictum_w. The type *rationalising explanation* admits of two sub-types for Hyman: there's the type of rationalising explanation that involves an explanatory link individuated by the *being guided by* relation, and there's the kind which involves a link individuated by the *treating as a premise* relation. Each of these types is a species of rationalising explanation: the *being guided by* type of explanation and the *treating as a premise* type of explanation are two different ways in which a rationalising explanation can be constituted. As a result, if Hyman is right then Williams' Dictum_w is false. That's because Hyman's view implies

that the explanation to which Smith's action is subject is of a type which requires p to be a reason in favour of the subject's ϕ -ing. This in turn implies that the particular explanation to which Smith's action is subject can be true only if p is a normative reason for S to ϕ . So Williams' Dictum_s is false too, if Hyman is right.

Putting all this together, Hyman's view in *Acting in Ignorance* seems to be this. ϕ -ing for the reason that p is a condition which obtains in both the good case and the bad. Moreover, it is a condition which consists in the obtaining of a rationalising explanation. However, rationalising explanations come in two forms: the being-guided-by-the-fact explanations and the treating- p -as-a-premise explanations. ϕ -ing for the reason that p consists in the former in the good case but in the bad case it consists in the latter.

I think that Hyman's view is essentially right, but the view which I want to defend departs from his in certain respects. Most significantly, I reject Hyman's claim that ϕ -ing for the reason that p is a condition which obtains in the bad case as well as the good. I have argued that it obtains only in the good case. However, I agree that ϕ -ing for the reason that p simply consists in the obtaining of a kind of rationalising explanation. And I agree that rationalising explanation comes in two fundamentally different forms. First, there's the neutral type which consists in an explanatory link between the fact of the subject's ϕ -ing and the fact of the subject's belief that p . This type of explanation is the only type which holds in the bad case and it does not suffice for one to count as ϕ -ing for a reason: one does no such thing in the bad case. Second, there's a type of rationalising explanation which takes the fact that p *qua* normative reason as *explanans* and takes the fact of the subject's ϕ -ing as *explanandum*. This second type of rationalising explanation holds in the good case only and the holding of it is indeed identical to what it is to respond to the normative reason that p . The former type requires that one believes that p and that one takes p to be a normative reason. In keeping with the account of the good case developed in §5.1, I want to say that the latter type of explanation is enabled by one's knowing that p and by one's knowledgeably taking p to be a normative reason for one to ϕ .

Unlike Hyman, I remain neutral on the nature of the concrete relations which ground the availability of each type of rationalising explanation. Moreover, unlike Hyman I make the claim that the good case explanation takes the fact that p as *explanans qua* normative reason. The thought is that the type of explanation which holds in the good case requires of its exemplars that they take the fact that p as *explanans*, but also that that fact is doing its explanatory work in so far as it is a normative reason. It is that feature of the fact that p which is its explanatory relevant property, when it comes to the sorts of good case rationalising explanations of which it is a part. Finally, like Hyman I build it into my position that the good case explanation is enabled partly by a state of knowing that p . But unlike Hyman I add to that that it is enabled also by a state of knowledgeably taking p to be a reason in favour of one's ϕ -ing as well.

The centre-piece of this Hyman-inspired position is the following thesis:

The Essential Normativity Thesis. There is a type of rationalising explanation which involves S's ϕ -ing being explained by the fact that p *qua* normative reason for S to ϕ , is enabled by S's knowledge that p and is enabled by S knowledgeably taking p to be reason for them to ϕ .

This encapsulates the thought that there is a special species of rationalising explanation available which takes the shape of a type of rationalising explanation that has the fact that p *qua* normative reason as *explanans*, enabled by one's knowledge. Explanations of this type require the presence of genuine normative reasons for the ϕ -ing in question, by dint of the very form of intelligibility they operate with.¹⁷ I'll refer to this thesis in abbreviated form as (ENT). (ENT) is my way of rejecting Williams's Dicta and hence Reductionism.

It is worth emphasising that what has been presented here is a certain thesis about rationalising explanation, (ENT), but, on top of that, I have also offered a positive conception of rationalising explanation which is made available by (ENT). I will be primarily focused on (ENT) itself in what follows. But in order to clarify my position, I want to make some points about (ENT) and its relation to the other elements of my positive picture.

The claim I make is that there is a special form of rationalising explanation to which the agent's ϕ -ing can be subject. The form in question places a restriction on the *explanans* of the explanations which exemplify it: that it must be a normative reason for the subject to ϕ which functions as the *explanans*, and it does so *qua* normative reason for them to ϕ . That is (ENT). I also suggest a number of other things. Chief amongst them is an account of the success condition of responding to the normative reason that p: that we should identify ϕ -ing in response to the normative reason that p with ϕ -ing subject to a rationalising explanation of the special sort described. What is the relation between (ENT) and that account of ϕ -ing in response to a normative reason?

(ENT) implies that there is a type of explanation which is present if, and only if, the subject is in the good case. The identity claim is a positive element of the overall position I argue for, which is made available by (ENT) given that it entails the bi-conditional just specified. Why does (ENT) entail that bi-conditional? That is: supposing there is a type of explanation which makes judgement sensitive phenomena intelligible essentially by appeal to normative reasons as such, why should it be that explanations of that type hold when, and only when, the subject is in the good case? Well, there are bad cases in which p is not a reason for the subject to ϕ . In those cases, the type of explanation envisaged by (ENT) doesn't hold, trivially. With respect to bad cases in which p happens to be a normative

¹⁷It is worth offering the reminder that I am not here suggesting that a *complete* rationalising explanation is one which takes the fact that p *qua* reason as *explanans*. Nor do I suggest that a complete neutral type of rationalising explanation takes the fact that one believes that p as *explanans*. Where ϕ -ing is an action, for example, what one represents to be the good in the action one is pursuing might also be a part of the former, and the state which represents that good might be a part of the latter, as might many other elements, for all I've said. I'm here bracketing off all elements of the explanation which are not contributed by the agent's cognitive life.

reason for the subject to ϕ , the subject still blamelessly fails to respond to a normative reason. I do not see how this is compatible with them ϕ -ing subject to a rationalising explanation of a form which makes their ϕ -ing intelligible by appeal to the normative reason as such. The point of saying that the agent fails to respond to what so happens to be a reason for them to ϕ is precisely that the reason doesn't figure in the explanation of their ϕ -ing, or if it does so then that is an accident. The upshot is that the sort of explanation described by (ENT) applies to the subject's ϕ -ing *only if* the subject is in the good case. But if (ENT) is true, then presumably the type of explanation is instanced somewhere, and where else could be instanced other than in good cases, given what has been said about the impossibility of it holding in bad cases. Thus, if (ENT) is right, then S ϕ s in response to the normative reason that p if, and only if S's ϕ -ing is subject to a rationalising explanation of the form envisaged by (ENT). This puts me in a position to endorse the identity claim: ϕ -ing in response to a normative reason consists in the obtaining of an explanation type envisaged by (ENT).

In addition to the account of the good case which is made available by (ENT), a further element of my positive position is the claim that there is a type of rationalising explanation available that is neutral in the way Williams' describes. My view is not that the only type of rationalising explanation there is is the special type described by (ENT). I agree full well that the agent's ϕ -ing is subject to a rationalising explanation in the bad case, where that bad case explanation is also present in the good case, on top of the special explanation envisaged by (ENT). My view about the neutral explanation is that it takes the fact that one believes that p as *explanans*. It does not, *contra* Dancy, take what one believes as *explanans*, for that would involve denying the factivity of explanation why.

So we have two sorts of rationalising explanation: a neutral sort and an essentially successful sort. My account of rationalising explanation can thus be stated like this: rationalising explanation consists either in ϕ -ing because p *qua* normative reason, so that one ϕ s in response to the normative reason that p, or it consists in ϕ -ing because one believes that p, so that one ϕ s in a way that manifests one's taking p to be a normative reason. The account I offer is thus a *disjunctive* or *pluralistic* account of the form *rationalising explanation*. Non-psychologistic rationalising statements report the essentially successful explanations, psychologistic rationalising statements report the neutral explanations.

I want to end this section by exploring a way of attacking Williams' Dictum_w which is distinct from my own. My way of rejecting Williams' Dicta involves first, rejecting Williams' Dictum_w, which in turn motivates a rejection of Williams' Dictum_s. Rejecting Williams' Dictum_w involves arguing that there is a kind of rationalising explanation which requires the presence of a genuine normative reason. During a discussion in which he rightly accuses Dancy of subscribing to the claim that rationalising explanations are of a type which is neutral on the truth of p, McDowell protests that in addition to such a form of rationalising explanation there is also a form to which the fact that p is essential. That is the

generic idea associated with the rejection of Williams' Dictum_w. Here is how McDowell sums up his objection to Dancy:

There is no need to deny that an explanation in one of the forms Dancy countenances can provide some understanding of an action. But if we have only that understanding, we do not yet know an answer to a question that should concern us if we are interested in how the action manifests the agent's practical rationality at work. We do not yet know, and we ought to want to know, whether the action can be understood as a rational response to the fact in question. If it can, we can have an understanding of the action to which its being a fact that the agent is acting in the light of is integral. That is the idea of acting in the light of a fact that is missing from Dancy's thinking. (McDowell, 2013a:19)

I reject Williams' Dicta and hence agree that there is a kind of intelligibility special to the good case, to which the truth (and indeed normative status) of *p* is integral. McDowell is quite right that that sort of intelligibility is available in the good case on top of the sort of neutral intelligibility Dancy countenances. But that is a generic idea. What is distinctive about the way in which I reject Williams' Dictum_w is the way in which I cash out the generic idea McDowell expresses. In my view, the special explanation is an explanation which appeals to the fact that *p* *qua* normative reason. Its *explanans* is *p* and *p* functions as the *explanans* by dint of its being a normative reason. It is enabled by one's knowledge that *p* and by one's knowledge of its normative status. Those features are integral to the type of rationalising explanation special to the good case.

However, there are other ways of conceiving of the sort of explanation special to the good case – there are other ways of making sense of the generic idea of the special kind of intelligibility to which McDowell refers. One alternative way of conceiving of it can be extracted from the work of Hornsby. Hornsby thinks that if one is married to a conception of rationalising explanation according to which rationalising explanation always takes the form of ϕ -ing because one believes that *p*, then one will not be able to make sense of the idea that an agent ϕ s in response to a normative reason. Instead, Hornsby recommends the following view:

(DisA) If *X* ϕ -d because *X* believed that *p*, then

EITHER *X* ϕ -d because *X* knew that *p*, so that *X* ϕ -d because *p*

OR *X* ϕ -d because *X* merely believed that *p*

(Hornsby, 2008:252)

Here is how I think we should read Hornsby. Hornsby takes it that ϕ -ing for a reason consists in ϕ -ing in a way that is subject to a rationalising explanation. One ϕ s in a way

that is subject to a rationalising explanation just in case one ϕ s because one believes that p. But ϕ -ing because one believes that p is a form of explanation which admits of two sub-forms. On the one hand, it can consist in ϕ -ing because one knows that p, which is equivalent to ϕ -ing because p. On the other hand it can consist in ϕ -ing merely because one believes that p: the *explanans* of the explanation could simply be one's belief that p, not one's knowledge or the fact that p. The first form of explanation is what ϕ -ing in response to a normative reason consists in.

One interpretive issue here is how Hornsby conceives of the explanation in the good case. She thinks that the explanations in terms of the agent's knowing that p are equivalent in some sense to the explanations in terms of the fact that p. Both explanations are non-neutral: they both require that p is true and hence that the subject is in the good case. But the former achieves that result partly by dint of taking a certain state of mind as *explanans*, a state of mind which guarantees that p is true. The second achieves that result partly by dint of taking the fact that p itself as *explanans*, although it is still an explanation that requires the subject to know that p. The second picture ascribes to the state of knowing that p the role of enabling condition for the special kind of good-case explanation. The first picture says that it is the *explanans* of the explanation of the special kind.

I do not propose to resolve the interpretive issue.¹⁸ All I want to point out is that there is an alternative to my proposal as to how to cash out the generic idea of a kind of rational intelligibility that is special to the good case which can at least be extracted from Hornsby's work, even if it cannot be attributed to her with confidence. When attempting to prove (ENT), I shall have to be careful, then, to show that my argument really does establish (ENT) and not the Hornsby-style alternative.

Having put on the table my favoured interpretation of Williams' Dictum, what its relationship is with Reductionism, and my way of resisting it – (ENT) – I now want to turn to the task of arguing for (ENT). The next section explores and rejects two extant attempts to undermine Williams' Dictum_w. This generates a need for a fresh argument, which I go on to supply.

5.5 Two Arguments against Williams' Dicta Rejected

This section examines two arguments which are supposed to tell against Williams' Dictum_w and finds those arguments lacking. §5.5.1 focuses on an argument from the Epistemic Thesis which is endorsed by Hornsby (2007a, 2007b, 2008) and McDowell (2013a). §5.5.2 focuses on an argument from Roessler (2014). These are the central arguments in the liter-

¹⁸I suspect Hornsby is happy to allow that instances of 'S ϕ s because they know that p' and corresponding instances of 'S ϕ s because p' express the same explanation and indeed an explanation of the same type, but she sees no need to decide whether the *explanans* is p or that one knows that p. The important point is just that there is a type rationalising explanation which essentially requires knowing. If that is her view then the arguments for (ENT) to come later do show that the issue needs settling.

ature for the abandonment of the Williamsian picture. Given that they fail, that constitutes something of a problem for those wishing to abandon the picture. In the end, I will attempt to remedy the problem by supplying some fresh considerations for its abandonment in favour of (ENT).¹⁹

5.5.1 The Appeal to the Epistemic Thesis

There is a generic idea associated with the rejection of Williams' Dictum_w and then there is my specific way of cashing it out. The generic idea is that there is a form of explanation which being a rationalising explanation can consist in that makes essential reference to a condition that obtains only if p is a normative reason for S to ϕ . The specific idea is (ENT): that there is a form of rationalising explanation which appeals to the fact that p *qua* normative reason. An alternative way of cashing out the generic idea, inspired by Hornsby but not attributable to her with confidence, is that there is a form of explanation which essentially takes the agent's knowing that p as *explanans*.

Hornsby (2007a, 2007b, 2008) develops an argument in favour of her preferred conception of rationalising explanation, encapsulated in the quote from her supplied in §5.4, which appeals to the Epistemic Thesis. Her preferred conception of rationalising explanation is intended by her to simply fall out of the Epistemic Thesis: she moves, in her writings on the matter, directly from a, in my view successful, proof of the Epistemic Thesis²⁰ to her favoured conception of rationalising explanation. However we interpret Hornsby's view – whether we read it as a commitment to (ENT), a commitment to the knowledge-based alternative to (ENT) which can be extracted from Hornsby's work itself, or as neutral on which of those is correct – it is surely intended to constitute a rejection of Williams' Dictum_w in favour of the generic idea that there is a form of rationalising explanation which is essentially normative reason-involving. What Hornsby commits herself to, even if she doesn't conceive of things in this way, is thus the soundness of an inference from the Epistemic Thesis to the generic idea that encapsulates a rejection of Williams' Dictum_w. Is that inference sound?

I accept the Epistemic Thesis. So if there is some fault with the inference it must lie with its validity, given that the sole premise of the inference is the Epistemic Thesis. Indeed, I think the argument is a *non sequitur*: just because ϕ -ing in response to the normative reason that p requires knowing that p it does not follow that there is a form of rationalising explanation which consists in making the agent's ϕ -ing intelligible by appeal to the agent's knowledge (whether the state of knowledge or what's known).

¹⁹There is, I think, a further argument to be found in Roessler's paper against Williams' Dictum_w which appeals to the plausible thought that the agent can always tell why a certain rationalising explanation holds, whenever they know the explanation to be true in the first place. I do not have space to discuss that argument here, but, like the two arguments I do discuss, I do not think it is effective.

²⁰Or strictly speaking, the instance of it applied to bodily action.

To see why we can look again at the Davidsonian position. Davidson should acknowledge that knowing that p is required for ϕ -ing in response to the normative reason that p . Also, he *can* acknowledge it, consistently with his claim that the form of rationalising explanation always involves appealing to the agent's causally active state of believing. For he can just say that even though the form of explanation to which rationalising explanations belong takes the shape just described, still, in the good case particular explanations appeal to states of belief which just are states of knowing. The agent's knowledge that p thus is the *explanans* in the good case, even though the form of explanation doesn't require it to be, so that it is not the *explanans qua* knowing. That is how Davidson might acknowledge the Epistemic Thesis consistently with his acceptance of Williams' Dictum_w, and indeed Williams' Dictum_s, depending on how he thinks of knowledge.²¹

The coherence of this position proves that the Epistemic Thesis all by itself does not entail the falsehood of Williams' Dictum_w. To think that it does is to confuse the claim that a certain condition is required for the success state of responding to the normative reason that p with the stronger claim that it is part of the way in which the explanation makes things intelligible that it appeals to that condition – that the condition in question is part of the form of the explanation. The inference to which Hornsby is committed to thinking sound is thus invalid.

There is a possible response to this on Hornsby's part which I now want to address. Davidson can pull off the trick of allowing the state of knowing an explanatory role in the good case because he *identifies explanantia* of rationalising explanations with states of mind.²² But, it might be pretested, this not the right thing to say about the ontology of *explanantia* – a topic which I have been remaining neutral on so far. The correct thing to say is that *explanantia* (and indeed *explananda*) are *truths*: true propositions which constitute modes of presentation of the concrete particulars the explanation concerns.²³ With this thesis about explanation in tow, it will be suggested that the Davidsonian will need to drop their claim that causally active states of mind are the *explanantia* of rationalising explanations in favour of the claim that truths which record the presence of such causally active states of mind are the relevant *explanantia*. Thus, it is not Smith's state of believing that there is free coffee available in the quad which strictly speaking explains why he walks to the quad but *that he believes that there is free coffee available in the quad* which does so. But with that conception of rationalising explanation in place, it is no longer possible for

²¹The same point could've been made by appeal to the development of Dancy's theory described above.

²²Again, it is worth emphasising that I am following the orthodox interpretation of Davidson's position here. It must be acknowledged, however, that Davidson himself distinguishes between the causes of an event – which are thought of as concrete particulars – from statements about such causes (Davidson, 1967). It is the latter which are identified by him with the *explanantia* of explanations. Thus, the conception of explanation which I'm about to describe and which enables Hornsby to sidestep the worry is actually more authentically Davidsonian than the conception of explanation operated with by the thesis I've labelled 'Davidsonian' here.

²³This is the conception of explanation operated with by Strawson (1985).

the Davidsonian to incorporate the Epistemic Thesis by making the move of identifying the *explanans* of rationalising explanations in the good case with a state of knowing, for the *explanans* is no longer thought of as a *state* in the first place.

The problem with this response is that the proponent of the Davidsonian position can accept the claim in the ontology of explanation and accept the thought that that claim precludes them absorbing the truth of the Epistemic Thesis *in the way described above*, without accepting that it disables them from absorbing the thesis altogether. For they can just incorporate the Epistemic Thesis, within the new ontological framework, by accepting that the *explanans* of the particular rationalising explanations in the good case is the truth that the agent knows that *p*. It's just that they will have to say that it is not built into the very form of explanation that it takes truths about the agent's knowledge as *explanantia*, so that they will have to accept that the truth that the agent knows that *p* does not explain why the agent *φs* in so far as it is a truth which records the presence of a state of *knowing*, but in so far as it records the presence of a mental state which constitutes the appearance of normative reasons to the subject. Hornsby's inference from the Epistemic Thesis to the denial of Williams' Dictum_w still fails, then.

5.5.2 The Anscombean Argument

I now move on to studying an argument from Roessler (2014), which I call *The Anscombean Argument*. The Anscombean Argument is concerned with what sense the proponent of Williams' Dictum_w can make of a certain sort of conversational exchange involving the asking and answering of the sort of 'why did you *φ*?' question that is a topic of discussion in Anscombe (1957). Consider Roessler's example of such an exchange: a case in which a photographer is setting up his camera on the side of the pavement in order to take a photograph of Marilyn Monroe. The photographer is asked 'why are you setting up your camera here?' and he replies: 'because Marilyn Monroe is going to come this way'. Here is what Roessler wants to say about the relevant sort of conversational exchanges:

To locate the disagreement [with Williams' Dictum_w], it is useful to distinguish three elements of Anscombe's view of the reason-seeking question 'why?':

- (a) We ordinarily take it that a good way to answer the question 'Why are you doing A?' is to answer the question 'What's the point (or: What's the good) of doing A?'
- (b) While there is more than one style of answering the latter question, the most basic way to do so is this: we set out the considerations in the light of which our doing A can be seen to be an effective way to promote some (in some way) desirable outcome.

- (c) Such explanations are often meant to be taken at face value. For example, accepting the reason given by the photographer for positioning his camera on that pavement would involve accepting (i) that Marilyn Monroe is going to pass by, and (ii) that this (plus the desirability of his taking a picture of her) gives him good reason for acting, and (iii) that he is acting because these considerations give him good reason.

(Roessler, 2014:5)

Anscombe claims, in other words, that one way of successfully answering her ‘Why?’ question is to offer an explanatory sentence which provides an answer to the question: ‘What’s the point of doing A?’, where that in turn consists in citing the normative reasons one is committed to their being as *explanantia*. The result is a non-psychologistic rationalising statement associated with the good case: S is ϕ -ing because p. Moreover, she thinks that we often intend such non-psychologistic rationalising statements to be taken as they appear: as purporting to explain just by appeal to the normative facts. Ways of doubting the explanation, then, would include doubting that p is true, that it has the normative significance being attributed to it, and that it is the factor doing the explanatory work – instead of some other, perhaps non-rational factor.

The problems for Williams Dictum_w are supposed to begin to emerge once it’s pointed out that its proponent is committed to a certain sort of interpretation of such conversational exchanges. The proponent of that view must, it seems, deny (c), on pain of convicting ordinary thought and judgement about reasons explanations of systematic error. To unpack this thought somewhat, the idea is that the proponent of Williams’ Dictum_w is committed to denying that rationalising explanations can take normative reasons as *explanantia*. But if (c) is correct, then we are pre-philosophically committed to the thought that they can. So if (c) is correct then the proponent of Williams’ Dictum_w would be committed to saddling ordinary thought and talk about reasons with systematic error. That would serve to shift the burden of proof onto the proponent of Williams’ Dictum_w, which is a result that they will presumably wish to avoid (wisely, given that, as we will see, there aren’t any good arguments for the thesis). So the proponent of Williams’ Dictum_w must resist (c): we are not pre-philosophically committed to the possibility of rationalising explanations that appeal to normative reasons, they must say.

The next stage of Roessler’s argument purports to establish that (c) is true. This is what Roessler has to say about the matter:

The case for (c) turns on the thought that practical reasoning has two aims that are intelligibly and, for the deliberator, essentially connected. One aim is to *get right* what one has most reason to do. The other aim is to *determine* or control what one will be (or is) doing. (*Ibid.*)

I think the correct way to understand what's going on here is as follows. Reasoning or deliberation which characteristically results in one's ϕ -ing for a reason involves two aims. The first is to bring to mind the set of facts which are reasons for and against one's ϕ -ing and to settle the issue of whether those facts, all things considered, speak in favour of one's ϕ -ing or not: this is a matter of working out what one has most reason to do. The second is to result in a ϕ -ing which is in some way controlled or determined by one's episode of reasoning. To say that these are *aims* of reasoning is to say there the episode of reasoning counts as defective to the extent that it fails to achieve either of them. Moreover, they are aims of reasoning *from the subject's own point of view*: from the subject's point of view, what they are doing is attempting to settle the matter about whether or not to ϕ and then to ϕ in a way that is in some sense controlled by their episode of deliberation.

The key thought is that from the subject's own point of view, the two aims are connected:

The two things are evidently connected. What is essential for success in practical reasoning is not just that it should get things right and that it should somehow make a difference to what one will eventually be doing, but that the facts (which one needs to get right) will make a difference to what one will be doing. (Roessler, 2014:5)

The thought is that one doesn't simply aim to get right whether one ought to ϕ and *in addition*, or *over and above that*, aim at a ϕ -ing controlled by one's deliberation. Rather, there is a single aim of reasoning which subsumes both aims already mentioned and links them: reasoning which characteristically results in a ϕ -ing for reasons is reasoning aimed at *one's ϕ -ing in a way that is explained by the reasons which settle the issue*. And this thought about what links the aim of determining what can be said in favour of one's ϕ -ing with the aim of determining what one does is supposed to tell in favour of (c):

To insist that the correct explanation of the agent's doing A lies in her 'motivating reasons' (her non-factive attitudes)... One would have to think about one's action from a standpoint that is neutral on whether one is getting things right in one's practical reasoning... It may not be *impossible* for an agent to adopt that kind of perspective even in the midst of deliberating and acting, but doing so would certainly go against the grain. It is hard to stop caring about the distinction between discovery and delusion while being engaged in a project of practical reasoning, the success of which turns on that distinction. (*Ibid.*:6-7)

In a word, we don't pre-philosophically conceive of the episodes of reasoning which result in our ϕ -ing for reasons as episodes which aim to deliver us ϕ -ings that are explained by the appearance of normative reasons, but ϕ -ings that are explained by the normative

reasons themselves. That is the point of the thought about what reasoning aims at from the agent's own point of view described above: the two aims described are united and subsumed by a more general aim, which is part of the subject's own point of view on their reasoning, that one's ϕ -ing is explained by the normative reasons one has surmised during the course of the reasoning.

The Anscombean Argument has two steps. The first aims to establish that (c) generates a problem for the proponent of Williams' Dictum_w. The second aims to establish that (c) is true. The problem, it seems to me, lies with the first step. The argument that (c) is incompatible with Williams' Dictum_w appeals to the idea that the proponent of it is committed to saying that rationalising explanations never take normative reason-involving *explanantia*. But, as we have seen, that is false: Davidson and Dancy can perfectly well agree that the *explanans* in the good case is either a state of knowing or the fact that p. So even if we are pre-philosophically committed to the claim that normative reasons explain why we ϕ , that's consistent with Williams' Dictum_w. It might even be consistent with Williams' Dictum_s, depending on how the particular Reductionist theory is developed.

5.6 Against Normative Epiphenomenalism

The state of play, dialectically speaking, is this. I've distinguished particular explanations from the form or type they exemplify. This helped us to distinguish between two ways of understanding Williams' Dictum. On one understanding of that thesis, it is equivalent to my target: Reductionism about ϕ -ing for the reason that p. My tactic is to refute Reductionism by refuting Williams' Dictum on that reading. I've surveyed some of the existent arguments against it and they have been found wanting, in light of the distinction between particular explanations and forms of explanation. This generates the need for some fresh considerations. That is what I attempt to provide here. In §5.6.1 I attempt to demonstrate that (ENT) should be our default view in this area. In §5.6.2 I attempt to show that there are no good arguments for Williams' Dictum_w. I conclude that we should adopt (ENT).

5.6.1 Shifting the Burden of Proof

It is a fundamental commitment of the proponent of Williams' Dictum_w, however their favoured theory is developed, that rationalising explanations which appeal to normative reasons or our knowledge of them never appeal to them *qua* normative reasons or *qua* knowledge. It is not by dint of p being a normative reason that it explains why one ϕ s, if it does, and likewise it is not by dint of its being a state of knowledge that a state of knowledge explains why one ϕ s either. In the first case, it is the fact that the reason is an apparent normative reason that does the explanatory work. In the second case it will be because the state in question constitutes the appearance of normativity to one that it does

the explanatory work. Whatever the positive story, the fact remains: if a normative reason-involving condition does explanatory work, for the proponent of Williams' Dictum_w, it is not by dint of being a normative reason-involving condition in the way it is that it does so. That is a result of Williams' Dictum_w because according to it the way in which explanations of the rationalising form make things intelligible is not by appeal to conditions that are normative-reason involving, but only by appeal to conditions which involve the appearance of normativity. If a normative reason-involving condition plays the role of a rationaliser, then, that will be accidental from the way in which the explanation makes things intelligible.

Contrast that with (ENT). It is precisely by dint of being a normative reason that a given normative reason gets to do the explanatory work it does, for that is part of the way in which rationalising explanations of the form posited by (ENT) make judgement sensitive phenomena intelligible. We would not have an explanation of the same form on our hands – things would not be made intelligible for us in that very way – if the *explanans* of such an explanation were traded for a condition which doesn't involve normative reasons.

I think this is a problematic commitment of Williams' Dictum_w – it commits it effectively to the thought that the normativity of normative reasons is an explanatorily epiphenomenal feature of them. But why is that problematic?

In Chapter One, I brought the phenomenon of responding to the normative reason that *p* into focus by presenting it as an analogue of the Kantian notion of ϕ -ing in the light of duty. One of the hallmarks of ϕ -ing in the light of duty is that it is a matter of ϕ -ing in response to a duty conferring feature *as such* or *qua* duty making feature. The phenomenon of responding to a reason displays a similar aspect: it is responding to a normative reason *as such* or *qua* normative reason.

This *as such* aspect of ϕ -ing in response to the normative reason that *p* is what distinguishes the sense of responding to reasons I'm interested in from the sense in which, for example, someone who is on a diet can be said to be responding to a normative reason in eating a slice of birthday cake they know they shouldn't. That it is their birthday is a reason for them to eat a slice and they are acting in a way that can be said to be a response to that fact when they engage in their akratic action. This is so even though they know full-well that they have decisive reason not to take a slice. This, I think, is not an example of the sort of phenomenon I'm interested in because the akratic agent is not responding to the fact that it is their birthday *in so far as it is a reason*, for they precisely acknowledge that the normative force of the reason fails to make a decisive case for action.

Now, this *as such* aspect of the success notion raises a challenge to the proponent of Williams' Dictum_w. For let us ask the question: what does it amount to other than the fact that the normative dimension of the normative reason to which the agent is responding does explanatory work with respect to the agent's ϕ -ing? What, that is, can ϕ -ing in response to a reason *qua reason* be, other than ϕ -ing in a way that is subject to an explanation which

depends on the normative dimension of the reason in question? Without a convincing response to this challenge we will have to conclude that normative reasons do explanatory work in so far as they are normative reasons, *contra* Williams' Dictum_w.

However, the proponent of Williams' Dictum_w does seem to have a response to this challenge, for there is a sense in which the agent in the bad case ϕ s because of their belief that p *in so far as p is a reason*. We need to acknowledge that there is a sense in which that's so in order to get the desired contrast between bad cases and cases of non-rational ϕ -ing, including cases of weakness of will, on the one hand, and the desired similarity with good cases of ϕ -ing for reasons, on the other hand. But it is agreed on all sides that the subject does not ϕ because p *qua* normative reason in the bad case. So the proponent of Williams' Dictum_w can say, in response to the challenge, that whatever ϕ -ing for a reason *qua* reason amounts to, it cannot amount to ϕ -ing in a way that is explained by a normative reason as such, for the *qua* reason aspect appears in the bad case too, where no such explanation is possible.

In response to this challenge I want to suggest that the *qua* reason aspect surfaces at the level of the subject's own point of view in a way that suffices for the subject to be committed to thinking of their ϕ -ing as explained by a normative reason *qua* normative reason. So even *if* the *qua* reason aspect at issue is present across good cases and bad, it can still be appealed to in order to shift the burden of proof onto the proponent of Williams' Dictum_w.

Contrast the subject who eats a slice of birthday cake in an akratic way with the subject who eats a slice in a non-akratic way. The first eats a slice in response to the fact that it is their birthday, but not *qua* reason, for they are on a diet and so are committed to thinking that the fact that it is their birthday is not a sufficient reason to eat the slice. The second eats a slice in response to the fact that it is their birthday *qua* reason. They are not on diet, and so that fact can function as a decisive reason for them. I submit that there is a difference between the way in which each subject conceives of their action: there is a difference between each action from the point of view of the agent who performs it. This difference is manifested in what the subject would be willing to say about their action. The akratic agent would be prepared to say things like: 'I know I shouldn't, but I don't care what's good for me: it's my birthday so I'll eat a slice of cake anyway!' whereas the non-akratic agent wouldn't be prepared to say such things.

What I want to suggest is that we should interpret the difference as partly a difference, from each agent's point of view, in how their respective actions are to be explained. The non-akratic agent thinks of their ϕ -ing as being explained by the normative reason that it is their birthday *qua* reason, whereas the akratic agent doesn't think of their action as being explained in that way. That is at least part of what the difference between an agent who ϕ s because p *qua* reason and an agent who doesn't but still counts as (non-rationally) responding to p consists in. The *qua* reason dimension of responding to normative reasons is

present from the point of view of the subject on their ϕ -ing *as* an explanatory phenomenon. It is the *qua* reason dimension which sets akratic ϕ -ings apart from ϕ -ings which are responses to reasons, and part of what the difference involves is a difference in how the ϕ -ing is to be explained from the point of view of the subject themselves.²⁴

It should now be clear that we can acknowledge the thought, if we like, that the *qua* reason dimension is exemplified in both good cases and bad, whilst also thinking that it causes problems for the proponent of Williams' Dictum_w. For even if that is so, and I haven't doubted it here, still: the *qua* reason dimension suffices for it to be the case that from the subject's own point of view (if not in fact), the subject's ϕ -ing is to be explained by *p qua* normative reason. The proponent of Williams' Dictum_w is thus committed to an error theory of our ordinary conception of ϕ -ing for reasons. This suffices to shift the burden of proof onto them.

I've raised a challenge to the proponent of Williams' Dictum_w that appeals to the *as such* dimension of responding to a normative reason. But the same challenge can be posed to the knowledge-based alternative to (ENT). According to that view, we should indeed reject Williams' Dictum_w, but we should not identify the *explanantia* of rationalising explanations with the fact that *p*. Instead, we should identify them with the state or fact of the agent's knowing that *p*. The problem for this view is that it too cannot, on the face of it, handle the *as such* dimension of responding to the normative reason that *p*. As I have said, that feature of the phenomenon we're interested in surfaces at the level of the subject's own point of view in a way that involves the subject being committed to their ϕ -ing being explained by the normative reason that *p* in so far as *p* is a normative reason. But according to the knowledge-based alternative to (ENT), the *explanans* isn't something which exemplifies the normative status of *p*, simply because the *explanans* isn't *p* but rather the agent's knowledge. So the way in which the *as such* dimension of the phenomenon we're interested in surfaces for the subject themselves also poses a difficulty for the knowledge-based alternative. The upshot of all of this is that (ENT) should be our default view.

5.6.2 Williams' Dicta are Undermotivated

Why might one endorse Williams' Dicta? I suspect that those who explicitly endorse them, like Dancy (2000) and Williams himself, take them to be obvious and not in need of argument. They are not obvious, however, and do need argument. In this section I examine and reject the best ways of arguing for them. Given that (ENT) should be our default view, the upshot will be that we are licensed in accepting it, as I will go on to detail.

I'll begin with what can be said in favour of Williams' Dictum_s. In fact, the Motivating

²⁴In Chapter One I contrasted responding to normative reasons, as well as bad cases thereof, not just with akratic ϕ -ings but with other sorts of non-rational ϕ -ings which could be described using similar language. Each of those non-rational phenomena fail to display the *qua* reason aspect, just as the akratic phenomena do. As a result, I could've just as easily relied on these other non-rational phenomena for making my point.

Reasons Postulate can be relied upon to motivate Williams' Dictum_s. It follows that every argument for the Motivating Reasons Postulate helps in the case for Williams' Dictum_s, and indeed arguing for the Postulate seems to me to be the only way to motivate Williams' Dictum_s. Chapter Three undermined the strongest arguments for the Postulate. Thus, the only way of arguing for Williams' Dictum_s has been undermined too.

This attempt at undermining what can be said in favour of Williams' Dictum_s hinges on the thought that the Motivating Reasons Postulate can be used to motivate it. But why should we think that? Well the Motivating Reasons Postulate identifies our ordinary notion of the reason which motivates the agent to ϕ with an item which either is, or is a part of, the *explanans* of the particular rationalising explanation to which one's ϕ -ing is subject just in case one ϕ s motivated by a reason. It then says that *that very explanation* holds in the bad case, thus delivering us the result that the agent ϕ s for the reason that p in the bad case too. Now if we add to this the premise that the only rationalising explanation which holds in the good case is the explanation which constitutes one's ϕ -ing for the reason that p, we get the result that the only rationalising explanations there are are normative-reason neutral, which is Williams' Dictum_s.

Of course, one might respond to this by saying that there are good reasons to think that there are rationalising explanations which apply in the good case which are not identical to the subject ϕ -ing for the reason that p.²⁵ However, what has emerged is that the argument goes wrong anyway, because the Postulate is undermotivated. The five arguments I have examined for the claim that S ϕ s for the reason that p in the bad case – Dancy's Master Argument, The Argument from Rationality, The Argument from Incredulity, The Argument from Transparency, and The Argument from Doubt – are supposed to be arguments for the Motivating Reasons Postulate. I rejected those arguments earlier. The refutation of the Motivating Reasons Postulate is thus revealed to be significant not just because it helps us to settle the Question of Psychologism but because it helps to settle the question of whether we should endorse Williams' Dictum_s.

I now move on to examining what can be said in favour of Williams' Dictum_w. One way of arguing for it is by appeal to Williams' Dictum_s, which entails it. This strategy, however, is no good because, as we have just seen, there are no good arguments for Williams' Dictum_s. The only other strategy for motivating Williams' Dictum_w that I know of consists in motivating Reductionism, and from that inferring Williams' Dictum_w, which is entailed by it. This, in turn, raises the question of how we might motivate Reductionism, independently of Williams' Dictum_w. I know of only one way of doing that, which I now want to detail and, in the end, undermine.

I've identified Reductionism with a certain identity claim. The claim is that ϕ -ing in response to the normative reason that p is identical to ϕ -ing subject to a rationalising explanation thought of as belonging to a type which is normative-reason neutral, in addition

²⁵This, I think, is what Hornsby would say.

to the obtaining of the set of extra-explanatory factors such as truth, normative status and knowledge which are individually necessary and, together with the holding of the particular explanation, sufficient for one to ϕ in response to the normative reason that p . Those factors, although extra to the type of explanation in question, will play a certain sort of explanatory role, as we have seen.

It is not built into the Reductionist's position that we can give a non-circular analysis of the concept of ϕ -ing in response to the normative reason that p . It might be that that notion can only be satisfactorily accounted for, ultimately, by appeal to some factor our concept of which can itself only be analysed partly by appeal to the concept of ϕ -ing in response to the normative reason that p . That would be consistent with Reductionism, which just says that the condition we have in mind when we think of an agent as ϕ -ing in response to the normative reason that p is identical to a condition that factors into ϕ -ing subject to a rationalising thought of in the way Williams' Dictum_w recommends, in addition to the relevant further factors. That metaphysical thesis is entirely consistent with the conceptual thesis that our analysis of the concept of ϕ -ing in response to the normative reason that p must ultimately be circular.

However, the thought that there is a satisfactory non-circular analysis of the concept of ϕ -ing in response to the normative reason that p can be appealed to in order to motivate Reductionism. For let's suppose that such an analysis is possible. Then it would follow that the concept of ϕ -ing in response to the normative reason that p is to be identified with a certain conjunctive concept: the concept that is the conjunction of the concepts corresponding to each of the conditions in the Reductionist analysis. Moreover, each of the concepts which compose the conjunctive concept would themselves be graspable independently of the concept of ϕ -ing in response to the normative reason that p . That is the shape that the non-circular analysis of the concept of ϕ -ing in response to the normative reason that p would have to take. But if that's so, then Reductionism would follow, for it would follow that what we have in mind when we think of an agent as ϕ -ing in response to the normative reason that p just is those set of conditions with which the Reductionist identifies ϕ -ing in response to a normative reason.

So the question now is this: what can be said in favour of a non-circular analysis of the concept of ϕ -ing in response to the normative reason that p ? I don't think there is any decisive reason to expect that such a thing is possible at the outset of inquiry. However, it might be thought plausible that the expectation that we can find a non-circular analysis of any given concept is something that we are justified in holding until it is proven that the concept needs to be taken as not admitting of one. Thus, we might reasonably expect to be able to find a non-circular analysis of the concept of ϕ -ing in response to the normative reason that p , until it is proven that such a project is impossible to complete. But with that point settled, we get a motivation for Reductionism and hence Williams' Dictum_w. For if we are entitled to expect a non-circular analysis of the concept of ϕ -ing in response to the

normative reason that p until it is proven that such a thing is impossible then, given that the possibility of such an analysis entails Reductionism and Reductionism entails Williams' Dictum_w, we are also entitled to take Williams' Dictum_w to be true until it is proven that a non-circular analysis of the success notion is not possible.

I do not think this argument is effective, for there are indeed decisive reasons to think that a non-circular analysis of the success notion is impossible. To see why, I want to begin with a feature of the type *rationalising explanation* to which Dancy (2004b) draws attention. We can start with what he says in the following passage, in which he is commenting on what he labels a 'Humean' conception of rationalising explanation – a conception of rationalising explanation which, we can take it, is the conception I've been associating with Davidson:

An intentional action, we are told, cannot occur in the absence of a pair of distinct states in the agent, one a belief one a desire, which states are to function as causes. But not just this: there is the further requirement that the belief and the desire be somehow 'appropriate' to each other; we must have an 'appropriate' pairing if we are to have the sort of explanation of action that the Humeans conceive themselves to be talking about. What this means is that an explanation of someone's taking the bus that appeals to his wanting to get to the market and believing that the bus is a convenient way of getting there is a good explanation of the Humean sort. A bad explanation of the Humean sort would be one that explained an action of putting on one's hat because one believes that it is raining and wants to wear a hat if it is not raining. The latter, we are told, is *no* explanation, because the belief and the desire are not appropriately related. (*Ibid.*:30-31)

What Dancy is drawing attention to here is an essential feature of rationalising explanations as such. The feature is that in order for a rationalising explanation to count as making things intelligible in the distinctive way it does, it must be that the *explanans* of the rationalising explanation links up with the judgement sensitive phenomenon in the right sort of way. It must be that the *explanans*, or its content, counts intelligibly as a reason in favour of the agent's ϕ -ing from their own point of view so that there is, in principle at least, a course of reasoning or deliberation open to them which would take them from the *explanans*, or its content, to the conclusion to ϕ . This is why, in Dancy's example, a belief that it is raining and a desire to wear a hat if it is *not* raining cannot be invoked in principle to explain in the rationalising way why the subject is wearing a hat: an agent being in those states does not intelligibly add up the agent seeing the action in a positive light.

The claim is not that states of mind such as those Dancy cites are incapable of explaining why the agent ϕ s at all. Rather, the point is that they are not capable of explaining if the explanation is supposed to be a rationalising explanation. That isn't merely a gener-

alisation about all rationalising explanations, but a point about the nature of the kind of intelligibility we seek when we seek rationalising explanations – it is a comment about the form of rationalising explanations.

This point about the link between *explanantia* and *explananda* in rationalising explanations itself calls out for explanation. What calls for explanation is the thought that rationalising explanations which take *explanantia* that link in the appropriate way to the relevant judgement sensitive phenomenon can be correct, whereas those which don't cannot be correct if they are supposed to be rationalising explanations at all. Dancy labels the first type of explanation 'A', the second 'B', and supplies what I think is a plausible explanation of the constraint on rationalising explanations at issue:

It seems to me that the obvious suggestion here is that our a priori preference for A-type explanations is grounded in... the fact that in 'making sense' of the agent they portray him as rational, as responding to rational norms, in a way that B-style explanations do not. (Dancy, 2004b:31-32)

The thought here is that ϕ -ing in a way that is subject to a rationalising explanation is constitutive of ϕ -ing in a way that counts as following a 'rational norm', and this is why an explanation is rationalising only if the *explanans* is connected to the relevant judgement sensitive phenomenon in the appropriate way. For this explanation to work, it must be that ϕ -ing in a way that constitutes following a rational norm suffices for the *explanans* of the explanation to stand in the appropriate link to the agent's ϕ -ing. This, of course, raises the questions: what is a rational norm and what is the rational norm which one is following when one ϕ s subject to a rationalising explanation?

A rational norm is a general principle which tells us under what conditions one ought to ϕ . We have already come across examples of rational norms before: the principles of structural rationality described in Chapter Three. Those principles in particular tell us that when the agent has a certain set of attitudes, they ought to ϕ , for example:

(RN1) If S believes that p and takes p to be an undefeated reason in favour of their ϕ -ing, then S ought to ϕ

This rational norm – a norm of structural rationality – tells us that the agent should ϕ if they believe they have an undefeated reason to do so. But rational norms needn't always take the shape of specifying attitudes and then telling us that the agent ought to ϕ , given those attitudes. Instead, the rational norms at issue might operate the other way round:

(RN2) If S ϕ s, then S ought to believe that p and take p to be an undefeated reason in favour of their ϕ -ing

This norm tells us, roughly, that an agent ought only ever ϕ if they take something to be an undefeated reason for them to do so.

Now, the rational norms we've just come across are, in a certain sense, *subjective*. They are subjective in the sense that they link claims about the agent's ϕ -ing with claims about the agents' *attitudes*, specifically: those attitudes which constitute the appearance of undefeated normative reasons to the subject. But, as Dancy points out, there are such things as *objective* rational norms. These norms are general principles which link the agent's ϕ -ing not with *apparent* normative reasons but with *genuine* normative reasons. Consider the following principle:

(RN3) If p is an undefeated reason for S to ϕ , then S ought to ϕ ²⁶

So we have a distinction between subjective and objective rational norms and the suggestion that in order to explain why the *explanantia* of rationalising explanations must be appropriately linked to the agent's ϕ -ing in the way described we have to think of rationalising explanations as constitutively involving the agent responding to rational norms: general principles which tell the agent what, from the point of view of rationality, to do. The question now is: should we identify the relevant rational norm with a subjective norm like (RN1) or (RN2) or with an objective norm such as (RN3)? Dancy himself argues for the latter and indeed suggests that (RN3) is itself the norm at issue:

I suggested above that in offering A-type explanations we portray our agent as responding to a norm of rationality. The first question should then be: if so, to which sort of norm is the agent to be thought of as responding, the objective sort or the subjective sort? But to this question, the answer seems clear. Agents who deliberate and act in the light of deliberations do not say to themselves 'Well, I believe that p and desire to ϕ if p ; so even if I am wrong in believing that p and wrong to desire to ϕ if p , still I should ϕ ...'. That is, agents do not deliberate on their own states of mind as input; they deliberate on and act in the light of how (as they see it) things are. So if the agent is responding to any principle or norm of rationality, it will be to an objective norm which tells him what he has reason to do in which situations, not to a subjective one at all. (Dancy, 2004b:33)

²⁶Dancy himself does not formulate principles (RN1)-(RN3) in the way I have. In my formulation, the 'ought' is, on the surface at least, attached to the consequent of the conditional. In the way Dancy formulates them, the 'ought' has scope over the whole conditional, so that (RN1) becomes: S ought:(if S takes p to be an undefeated reason for them to ϕ then S ϕ s) and (RN3) becomes: S ought:(if p is an undefeated reason for S to ϕ , then S ϕ s). I take it that this reflects Dancy's view that oughts of subjective rationality have wide-scope, given that they just are oughts of normative reason-hood and that he needs to avoid bootstrapping worries. But reading Dancy in that way is not straightforwardly correct, given that there is no such motivation for thinking that the 'ought' of (RN3) has wide-scope. Whatever Dancy's reason for ascribing wide-scope to the 'oughts' which appear in his formulations of the relevant principles of rationality, it would make no difference to the discussion here if we were to follow Dancy in formulating the principles in that way.

The thought here is that an agent cannot count as following (RN1) in so far as a rationalising explanation applies to them. That would require that the agent takes the facts about their states of mind involved in the antecedent of the instance of the norm in question into account during the episode of deliberation which leads to their ϕ -ing, and this is not something which agents who ϕ subject to a rationalising explanation always, or even typically, do. Typically the agent's eye is directed outwards onto the normative reasons in the world when they engage in reasoning or deliberation which leads up to their ϕ -ing. This means that the agent who ϕ s subject to a rationalising explanation must be ϕ -ing in a way that is guided by (RN3), our objective principle. The principle which guides the agent's ϕ -ing when they ϕ for a reason is that they ought to ϕ , if there is an undefeated normative reason present. Following (RN3) is part of what it is for a rationalising explanation to apply to one, for that is what explains why rationalising explanations essentially take *explanantia* which are appropriately linked to the relevant judgement sensitive phenomenon.

It is worth emphasising that this point about rationalising explanation is supposed to hold whatever one's views about rationalising explanations are. Even if one thinks that the only rationalising explanations there are are ones which are normative reason-neutral, it would still be the case that those rationalising explanations essentially involve the subject following (RN3), and that this is why those explanations make things intelligible only if an *explanans* is appealed to which is appropriately connected with the relevant judgement sensitive phenomenon. One point that that brings out is that even the rationalising explanations which hold in the bad case will be ones the holding of which constitutes the agent ϕ -ing in a way that constitutively involves them following (RN3). In some bad cases, of course, p isn't really a reason for S to ϕ , in which case the subject will not count as *successfully* adhering to the norm in question. But, as Dancy himself points out, that doesn't mean that S thereby counts as successfully following a different norm, for example, (RN1). Even in the bad case, the subject is still following (RN3), just unsuccessfully.

I think Dancy is right about all of this. However, I also think that it reveals that it is not possible to provide a non-circular analysis of the concept of responding to the normative reason that p . To see this, we can ask the question: what is it to *follow* (RN3)? We have a principle of rationality which tells one to ϕ , if there is undefeated reason for one to ϕ . It has been claimed that when an agent ϕ s subject to a rationalising explanation, they stand in a certain relation to that principle: that of *following it* or *adhering to it*. What is that relation?

Well, rational agent's are not, in general, concerned merely with ϕ -ing in line with the reasons there are, but are concerned with ϕ -ing in response to those reasons: they are concerned with ϕ -ing in the light of reasons as such. It is not as if we lead our normative lives attempting to perform actions or adopt attitudes which are, as it happens, acts and attitudes for which there are undefeated reasons, perhaps recognised as such by us. Rather, at least sometimes we show a concern for ϕ -ing in a way that stands in a certain relationship

with those reasons recognised as such: the responding to a normative reason relation. Responding to normative reasons, just like acting in the light of duty, is a status which rational agents as such manifest a concern for.

But now let us ask: what could it be say that an agent is *following* or *adhering to* (RN3) other than to say that they are manifesting their concern to respond to normative reasons as such? To say that an agent is following or adhering to the principle: if p is an undefeated reason for one to ϕ , then ϕ , just seems to be another way of saying: the agent is purporting to respond to a normative reason, or, alternatively put: the agent is manifesting a concern for making sure that they respond to reasons.

There is, of course, a further question about what this amounts to. But I do not need to go into that here. All I need to make my point is that the concept of responding to a normative reason needs to enter into our account of what it is to follow (RN3), and hence needs to enter into our account of rationalising explanation.

Properly developed, then, Dancy's point about rationalising explanation is that the 'because' of rationalising explanations requires the agent to count as following the objective norm (RN3). That is part of our understanding of what making things intelligible in the rationalising way is. But this should show us that it is not possible to provide a non-circular analysis of the concept of ϕ -ing in response to the normative reason that p . For let's suppose that we can give such an analysis. Then the analysis would have to appeal to our concept of ϕ -ing subject to a rationalising explanation. But what has emerged is that the concept of responding to the normative reason that p is one which we must invoke in order to understand the concept of rationalising explanation: our concept of rationalising explanation is a concept of a type of explanation which applies to agents only if they are following an objective norm, where following the norm in question is itself something that is itself a matter of manifesting a concern for ϕ -ing in response to normative reasons. The very concept of rationalising explanation is parasitic on the success notion. Thus, no non-circular analysis of the success notion is possible.

We can conclude then that the way of motivating Reductionism and hence Williams' Dictum_w which appeals to the possibility of a non-circular analysis of the concept of responding to the normative reason that p , fails. Since that is the only way of motivating Williams' Dictum_w other than by appeal to (the itself undermotivated) Williams' Dictum_s, we can conclude that Williams' Dictum_w is undermotivated.

I've already established that (ENT) should be our default view. So all I need to do in order to justify acceptance of (ENT) is to show that its opponents are undermotivated. I've just shown that its main opponents, Williams' Dicta, are undermotivated. That just leaves the knowledge-based alternative to (ENT), and with respect to that thesis it is not clear to me at all how one might motivate it, other than by appeal to the thought, already proven to be false, that it represents the only way to reject Williams' Dictum_w. We can conclude, then, that we're justified in accepting (ENT).

5.7 The Primacy of the Good Case

I have argued for a pluralistic account of rationalising explanation. Rationalising explanation can consist in a form of explanation which essentially appeals to normative reasons as such or it can consist in a form of explanation which appeals to the fact that the subject believes that p . The former type of explanation is present in the good case only, the latter across both good cases and bad. But putting it like this makes it explicit that in my view the subject's ϕ -ing in the good case is subject to two explanations, each of a different form: an explanation in terms of the normative reason to which they are responding and an explanation which appeals to their belief that p . This raises the question: what is the relationship between the two forms of explanation which each have instances in the good case?

Roessler and Hornsby both suggest that the neutral explanation is in a certain sense parasitic on the good case explanation. Here is how Roessler sums up the idea:

...it is not just that there is more than one way to explain intentional action 'as such', but we must acknowledge what is sometimes called the 'primacy of the good case'. The ability to make sense of actions in terms of non-factive rationalizing attitudes *depends* on the more basic ability to find action intelligible as something the agent has a normative reason to do (and is doing for that reason). (Roessler, 2014:8)

To say that the good case explanations are primary is to say that we are to understand what it is to provide a normative reason-independent explanation in terms of what it is to provide a normative reason-involving explanation. Our understanding of what it is to provide the first sort of explanation is parasitic on our understanding of what it is to provide an explanation of the second sort. Given the conception of rationalising explanation I have argued is correct here, the thought would be that giving a rationalising explanation in terms of the subject's believing that p needs to be understood by appeal to giving a rationalising explanation in terms of the normative reasons as such.

Hornsby and Roessler each offer what we might call *counterfactual theories* of the primacy of the good case. On this sort of account, the primacy of the good case is achieved via an analysis of the rationalising character of beliefs which appeals to some counterfactual that makes reference to the good case. Hornsby's thought is that we are to understand what it is to provide an explanation of an action in terms of belief by appeal to the idea that were what's believed true, it would be a normative reason for the agent's action. We then say, along with deniers of Williams' Dictum_w, that normative reasons as such have the power to rationalise ϕ -ings, and this gets us the result that explanatory potential of beliefs depends on that of normative reasons: it is because normative reasons have the power to

rationalise and the belief satisfies the counterfactual at issue that it in turn has the power to rationalise.²⁷

But as Roessler complains ‘why should counterfactual reasons be relevant to understanding the actual explanatory connection between beliefs and actions?’. The worry here is that although the explanatory potential of beliefs might be parasitic on that of normative reasons, it is not clear how this delivers what we want: that the neutral explanations cast in terms of belief are themselves parasitic on the normative reason-involving explanations. Roessler has an alternative counterfactual account which is supposed to fill in this lacuna:

We understand the explanatory role of mere beliefs on the model of that of knowledge. What merely believing the chocolate to be in the cupboard makes excusable is to act in ways that may in fact be pointless but that, if one’s belief were knowledge, would be intelligible in terms of a normative reason provided, in part, by the fact that the chocolate is in the cupboard. (Roessler, 2014:9)

The thought seems to be that to explain why the agent ϕ s by appeal to their belief that p is to provide an explanation which is true only if the following counterfactual holds: were S to know that p , they would ϕ in a way that is explained by the normative reason that p . Satisfying that counterfactual is part of what it is to be subject to a neutral rationalising explanation in terms of belief. Thus we get a parasiticness of the neutral explanation on the essentially successful explanation.

It seems to me, however, that Roessler’s account of the primacy of the good case is problematic. The problem is that there are possible cases in which the subject ϕ s because they believe that p , but where they don’t satisfy Roessler’s counterfactual. The cases in question are cases in which the subject comes to know that p in a way that would cause them to no longer ϕ in light of the apparent reason that p . Consider a case in which the subject ϕ s because they believe that p , but were the subject to come to know that p they would come to know it off the back of an informant who tells them that p is a good reason for them to ϕ . But let’s suppose further that our subject has a desire, for whatever reason, to rebel against the informant in question: what the informant recommends that they do, the agent will precisely not do. So the agent doesn’t satisfy the counterfactual: were S to know that p , they would ϕ in a way that’s made intelligible by their knowledge that p . Still, in the actual case where no informant is present, they ϕ because they believe that p .

At this point it might be possible for Roessler to finesse the counterfactual to which he appeals so as to avoid the worry. We would then have to check the fresh counterfactual for further counterexamples. I do not want to suggest that accounting for the primacy of the good case in counterfactual terms is doomed to fail. But I do want to point out that there is an alternative way of delivering the primacy of the good case. In fact, the

²⁷ See Hornsby (2008:258-259).

materials for constructing the alternative are already in place, for I argued earlier that part of what it is for an explanation to be a rationalising explanation is for it to involve the subject following the objective norm (RN3), where our account of what it is to follow that norm essentially makes reference to the notion of responding to a normative reason. I have also argued that the latter notion should be identified with the notion of ϕ -ing subject to a rationalising explanation of the essentially successful variety posited by (ENT). So we get the result that neutral rationalising explanations are to be understood in terms of the good case rationalising explanations. To say that S ϕ s because S believes that p is to say, partly, that in ϕ -ing, S is following the general principle: if p is an undefeated reason for one to ϕ , then ϕ , where following the latter needs to be understood in terms manifesting a concern to ϕ in a way that is subject to a type of explanation which requires the presence of normative reasons.

One question I have not addressed is what motivation there is for subscribing to the primacy thesis at issue in the first place. Its truth has indeed fallen out of the arguments of this chapter, and I've suggested that Roessler's way of cashing it out will at least need modifying, but none of this is particularly interesting if there is no reason to subscribe to the primacy of the good case in the first place. I do not want to settle the question of whether there is any need to subscribe to the primacy of the good case here. I rest content with the thought that *if* there is a good reason to subscribe to it, then that is no problem for my account, for it is easily accommodated.

Conclusion

Dancy thinks of his work in this area as constituting a response to Davidson and his followers. In particular, he thinks of it as a response to the psychologistic excesses of Davidson's view. The present work can be advertised as a response to both Davidson and Dancy (and indeed to the Neo-Davidsonians). In particular, it is a response to two excesses common to both Davidson and Dancy. We have the success case of responding to a normative reason as such. We also have bad cases of ϕ -ing in a way that would seem to one, were one to reflect on the matter, to be a case of responding to reasons but in which one blamelessly fails to achieve success. The first excess is a commitment to the claim that the subject ϕ s for the reason that p across good cases and bad, so that in so far as they count as ϕ -ing for the reason that p, they do not count as ϕ -ing in response to the *normative reason* that p. The second excess is a commitment to the claim that success is a matter of ϕ -ing motivated by a reason (in that special sense of 'reason'), plus additional further factors that hold over and above ϕ -ing motivated by a reason. Those excesses together constitute the Motivating Reasons View.

This view was introduced in Chapter Two, against the background of the basic material introduced in Chapter One. I argued against the Motivating Reasons View in Chapter Three by appeal to the suggestion that it should be our default view in this area that ϕ -ing for a reason just is the success phenomenon of responding to a normative reason. It was then argued that there are no good reasons to think that the subject ϕ s for a reason in bad cases, thus justifying our denial of the existence of motivating reasons in favour of the alternative just advertised.

Now, although this requires us to reject the thought that there are such things as motivating reasons in the special sense utilised by the Motivating Reasons View, it still leaves us with two issues: what position to take with respect to the Dialectic of Psychologism and whether we should agree that there is only one type of rationalising explanation, in terms of which we would need to offer a Reductionist account of success. Those issues are still live, it is just that they need to be re-formulated so that they don't involve a commitment to motivating reasons in the special sense.

I argued in Chapter Four that we should opt for an Anti-Psychologist view about the reasons for which we ϕ , and the refutation of the Motivating Reasons View has helped us undermine a powerful argument for it. Finally, I argued in Chapter Five that we must

acknowledge a special sort of rationalising explanation which essentially appeals to normative reasons and in terms of which we should account for our success notion. The theory which emerged in Chapter Five enables us to completely resolve the Dialectic of Psychologism by delivering us the result that when one ϕ s for a reason, the reason for which one ϕ s is identical to the *explanans* of the (or rather: a) rationalising explanation of one's ϕ -ing.

The positive position argued for here is a position which I advertised as *minimalist* in the introduction, for it relies, at most, only on two concepts of a reason: the concept of a normative reason and the concept of a reason why. It should now be clear what that view amounts to, and how it differs from the Motivating Reasons View. About the success case, I argue that we should identify it with ϕ -ing in a way that is subject to a special kind of explanation: an explanation the form of which requires the presence of normative reasons, which do their explanatory work *qua* normative reasons. The explanation is enabled by one's knowledge of one's reason and its normative status, so that one's ϕ -ing can be said to manifest such knowledge. About the latter, I say that it consists in ϕ -ing because one believes that *p*, in the rationalising sense of 'because'. That explanation is in play in the good case, but it is not the only explanation in play. Furthermore, it is parasitic on the essentially normative reason-involving explanation, for all rationalising explanations involve adhering to a rational norm which tell one to ϕ , if there is undefeated reason to do so, and adhering to such a norm is a matter of purporting to respond to a normative reason. I have also said that we should identify our ordinary talk of ϕ -ing for reasons with talk that reports the success phenomenon. It should be acknowledged, however, that there is some talk – 'S's reason' talk and Dancy's 'in the light of' talk – which, on closer inspection, can be read as recording the neutral condition.

There are a number of issues which would need to be explored in order to give a fuller treatment of the phenomena here. I haven't said anything, for example, about what it is to act in the pursuit of a goal, and how this relates to acting for reasons. I haven't said very much about what ϕ -ing in a way that is guided by a principle, for example a moral principle, is and how that relates to ϕ -ing for reasons. I haven't talked much about what the relationship is between ϕ -ing for reasons and the episodes of deliberation and reasoning which in some sense lead to one being in such states. I haven't said anything about the debate between internal and external reasons theorists and how this might affect our account of responding to reasons. I haven't investigated whether perceptual beliefs can be said to be held for reasons, in so far as they are based on perception. And a further vexed issue which has been unexplored is whether the rationalising 'because' is a causal 'because'. I intend to pursue these questions in future work. What has been provided here, I hope, is a framework within which those questions can be explored: a framework which differs markedly from that in which Davidson, Dancy and the Neo-Davidsonians operate.

Bibliography

- Alvarez, Maria (2010). *Kinds of Reasons*. Oxford: Oxford University Press.
- Amrstrong, D.M. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Anscombe, G.E.M. (1957). *Intention*. London: Harvard University Press.
- Bankhurs, David, Hooker, Brad, and Little, Margaret Olivia, eds. (2013). *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*. Oxford: Oxford University Press.
- Broome, John (1999). "Normative Requirements". *Ratio*, 12 (4): 398–419.
- (2004). "Reasons". In: *Reasons and Value: Themes from the Moral Philosophy of Joseph Raz*. Ed. by R. Jay Wallace, Philip Pettit, Samuel Sheffler, and Micahel Smith. Oxford: Oxford University Press: 28–55.
- (2007). "Wide or Narrow Scope?" *Mind*, 116 (462): 359–370.
- Collins, Arthur (1997). "The Psychological Reality of Reasons". *Ratio*, 10 (2): 108–123.
- Comesăna, Juan and McGrath, Matthew (2014). "Having False Reasons". In: *Epistemic Norms*. Ed. by JJohn Turri & Clayton Littlejohn. Oxford: Oxford University Press: 59–80.
- Dancy, Jonathan (1985). *Introduction to Contemporary Epistemology*. London: Blackwell.
- (2000). *Practical Reality*. Oxford: Oxford University Press.
- (2003). "Précis of *Practical Reality*". *Philosophy and Phenomenological Research*, 67 (2): 423–428.
- (2004a). *Ethics Without Principles*. Oxford: Oxford University Press.
- (2004b). "Two Ways of Explaining Actions". In: *Agency and Action*. Ed. by John Hyman & Helen Steward. Cambridge: Cambridge University Press: 25–42.
- (2008). "On How to Act – Disjunctively". In: *Disjunctivism: Perception, Action, Knowledge*. Ed. by Adrian Haddock & Fiona Macpherson. Oxford: Oxford University Press: 262–279.
- (2011). "Acting in Ignorance". *Frontiers of Philosophy in China*, 6 (3): 345–357.
- (2014). "On Knowing One's Reasons". In: *Epistemic Norms*. Ed. by JJohn Turri & Clayton Littlejohn. Oxford: Oxford University Press: 81–96.
- Davidson, Donald (1963). "Actions, Reasons, and Causes". *Journal of Philosophy*, 60 (23): 3–19. Reprinted in Davidson (2001): 3–20. All page references are to the reprint.

- Davidson, Donald (1967). "Causal Relations". *Journal of Philosophy*, 64 (21). Reprinted in Davidson (2001):149-162. All page references are to the reprint.: 691–703.
- (1969). "True to the Facts". *Journal of Philosophy*, 66 (21): 784–764.
- (2001). *Essays on Actions and Events*. Oxford: Oxford University Press.
- Davis, Wayne A. (2003). "Psychologism and Humeanism". *Philosophy and Phenomenological Research*, 67 (2): 452–459.
- (2005). "Reasons and Psychological Causes". *Philosophical Studies*, 122 (1): 51–101.
- Dodd, Julian (2008). *An Identity Theory of Truth*. Oxford: Oxford University Press.
- Frankfurt, Harry G. (1971). "Freedom of the Will and the Concept of a Person". *Journal of Philosophy*, 68 (1): 5–20. Reprinted in Frankfurt (1998):11-25. All page references are to the reprint.
- (1998). *The Importance of What we Care About*. Cambridge: Cambridge University Press.
- Gendler, Tamar and Hawthorne, John (2005). "The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies". *Philosophical Studies*, 124 (3): 331–352.
- Haddock, Adrian and Macpherson, Fiona, eds. (2008a). *Disjunctivism: Perception, Action, Knowledge*. Oxford: Oxford University Press.
- Hieronymi, Pamela (2011). "Reasons for Action". *Proceedings of the Aristotelian Society*, 111 (3): 407–427.
- Hooker, Brad and Struemer, Bart (2004). "Procedural and Substantive Practical Rationality". In: *The Oxford Handbook of Rationality*. Ed. by Peirs Rawling & Alfred R. Mele. Oxford: Oxford University Press: 57–74.
- Hornsby, Jenninfer (2007a). "Knowledge, Belief and Reasons for Acting". In: *Explaining the Mental*. Ed. by Carlo Penco, Michael Beany, and Massimiliano Vignolo. Newcastle: Cambridge Scholars Publishing: 88–105.
- (2007b). "Knowledge in Action". In: *Action in Context*. Ed. by Anton Leist. Berlin: Walter de Gruyter: 285–302.
- (2008). "A Disjunctive Conception of Acting for a Reason". In: *Disjunctivism: Perception, Action, Knowledge*. Ed. by Adrian Haddock and Fiona Macpherson. Oxford: Oxford University Press: 262–279.
- Hughes, Nick (2014). "Is Knowledge the Ability to ϕ for the Reason that p?" *Episteme*, 11 (4): 457–462.
- Hursthouse, Rosalind (1991). "Arational Actions". *Journal of Philosophy*, 88 (2): 57–68.
- Hyman, John (1999). "How Knowledge Works". *Philosophical Quarterly*, 50 (197): 433–451.
- (2006). "Knowledge and Evidence". *Mind*, 115 (460): 891–916.
- (2010). "The Road to Larissa". *Ratio*, 23 (4): 393–414.
- (2011). "Acting for Reasons: Reply to Dancy". *Frontiers of Philosophy in China*, 6 (3): 358–368.

- Kant, Immanuel (1996). *Practical Philosophy*. Ed. by Mary J. Gregor. Cambridge: Cambridge University Press.
- Kolodny, Niko (2005). "Why be Rational?" *Mind*, 114 (455): 509–563.
- (2007). "State or Process Requirements?" *Mind*, 116 (462): 371–385.
- Leist, Anton, ed. (2007). *Action in Context*. Berlin: Walter de Gruyter.
- Littlejohn, Clayton and Turri, John, eds. (2014). *Epistemic Norms*. Oxford: Oxford University Press.
- Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin.
- Marcus, Eric (2012). *Rational Causation*. Camb. Mass.: Harvard University Press.
- McDowell, John (1979). "Virtue and Reason". *The Monist*, 62 (3): 331–350. Reprinted in McDowell (1998a): 50–73. All page references are to the reprint.
- (1998a). *Mind, Value, and Reality*. Camb. Mass.: Harvard University Press.
- (2013a). "Acting in the Light of a Fact". In: *Thinking About Reasons*. Ed. by David Bankhurst, Brad Hooker, and Margaret Olivia Little. Oxford: Oxford University Press: 13–28.
- (2013b). "Perceptual Experience: Both Relational and Contentful". *European Journal of Philosophy*, 21 (1): 144–157.
- McHugh, Conor (2010). "Self-Knowledge and the KK Principle". *Synthese*, 173 (3): 231–257.
- Miller, Christian (2008). "Motivation in Agents". *Noûs*, 42 (2): 222–266.
- Olson, Jonas and Svensson, Frans (2005). "Regimenting Reasons". *Theoria*, 61 (3): 203–214.
- Parfit, Derek (1997). "Reasons and Motivation". *Proceedings of the Aristotelian Society*, suppl. vol. 71 (1): 99–130.
- Peacocke, Christopher (1998). "Conscious Attitudes, Attention, and Self-Knowledge". In: *Knowing Our Own minds*. Ed. by Crispin Wright, Barry C. Smith, and Cynthia MacDonald. Oxford: Oxford University Press: 63–98.
- Penco, Carlo, Beany, Michael, and Vignolo, Massimiliano, eds. (2007). *Explaining the Mental: Naturalist and Non-Naturalist Approaches to Mental Acts and Processes*. Newcastle: Cambridge Scholars Publishing.
- Pryor, James (2007). "Reasons and That-Clauses". *Philosophical Issues*, 17 (1): 217–244.
- Rawling, Peirs and Mele, Alfred R., eds. (2004). *The Oxford Handbook of Rationality*. Oxford: Oxford University Press.
- Raz, Joseph (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.
- Reisner, Andrew and Steglich-Petersen, Asbjørn, eds. (2011). *Reasons for Belief*. Oxford: Oxford University Press.
- Roessler, Johannes (2014). "Reason Explanation and the Second-Person Perspective". *Philosophical Explorations*, 3: 1–12.
- Ryle, Gilbert (2000). *The Concept of Mind*. Penguin.

- Sandis, Constantine (2011). *The Things We Do and Why We Do Them*. Palgrave Macmillan.
- (2013). “Can Action Explanation Ever be Non-Factive?” In: *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*. Ed. by David Bankhurst, Brad Hooker, and Margaret Olivia Little. Oxford: Oxford University Press: 29–49.
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Camb. Mass.: Harvard University Press.
- (2003). “Metaphysics and Morals”. *Proceedings of the Addresses of the American Philosophical Association*, 77: 7–22.
- Schroeder, Mark (2008). “Having Reasons”. *Philosophical Studies*, 139 (1): 57–71.
- (2011). “What does it take to “have” a reason?” In: *Reasons for Belief*. Ed. by Andrew Reisner and Asbjørn Steglich-Petersen. Oxford: Oxford University Press: 201–222.
- Setiya, Keiran (2011). “Reasons and Causes”. *European Journal of Philosophy*, 19 (1): 129–157.
- Smith, Michael (1987). “The Humean Theory of Motivation”. *Mind*, 96 (381): 36–61.
- (2003). “Internalism, Psychologism, and the Normative Story”. *Philosophy and Phenomenological Research*, 67 (2): 460–467.
- Steward, John Hyman & Helen, ed. (2004). *Agency and Action*. Cambridge: Cambridge University Press.
- Stout, Rowland (2004). “Internalising Practical Reason”. *Proceedings of the Aristotelian Society*, 104 (3): 229–243.
- (2009). “Was Sally’s Reason for Running from the Bear that She Thought it was Chasing Her?” In: *New Essays in the Explanation of Action*. Ed. by Constantine Sandis. Palgrave: Macmillan: 48–61.
- Strawson, P.F. (1985). “Causation and Explanation”. In: *Essays on Davidson: Actions and Events*. Ed. by Bruce Vermazen and Merrill B. Hintikka. Oxford: Oxford University Press: 115–135. Reprinted in Strawson (1992): 109–131. All page references are to the reprint.
- (1992). *Analysis and Metaphysics*. Oxford: Oxford University Press.
- Textor, Mark (2012). *States of Affairs*. Ed. by Edward N. Zalta. Stanford Encyclopedia of Philosophy. URL: <http://plato.stanford.edu/archives/sum2014/entries/states-of-affairs>. Accessed 1.08.2015.
- Turri, John (2009). “The Ontology of Epistemic Reasons”. *Noûs*, 43 (3): 490–512.
- (2011). “Believing For a Reason”. *Erkenntnis*, 74 (3): 393–397.
- Wallace, R. Jay (2003). “Explanation, Deliberation, and Reasons”. *Philosophy and Phenomenological Research*, 67 (2): 429–435.
- Wallace, R. Jay, Pettit, Philip, Sheffler, Samuel, and Smith, Michael, eds. (2004). *Reasons and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press.

- Williams, Bernard (1970). "Deciding to Believe". In: *Language, Belief and Metaphysics*. Ed. by H. Kiefer and M. Munitz. Albany: State University of New York Press. Reprinted in Williams (1973):136-151. All page references are to the reprint.
- (1973). *Problems of the Self*. Cambridge: Cambridge University Press.
- (1976). "Moral Luck". *Proceedings of the Aristotelian Society Supp. Vol. 50* (226): 115–135. Reprinted in Williams (1981): 20-39. All page references are to the reprint.
- (1980). "Internal and External Reasons". In: *Rational Action*. Ed. by Ross Harrison. Cambridge: Cambridge University Press. Reprinted in Williams (1981): 101-113. All page references are to the reprint.
- (1981). *Moral Luck*. Cambridge: Cambridge University Press.
- Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
- Wright, Crispin, Smith, Barry C., and Macdonald, Cynthia, eds. (1998). *Knowing Our Own Minds*. Oxford: Oxford University Press.